Queen Mary University of London

DEPARTMENT OF PHYSICS AND ASTRONOMY

Thesis submitted in fulfilment of the requirements
of the degree of Doctor of Philosophy

Cristiano Alpigiani

Search for Rare B Decays into Two Muons with the ATLAS Detector

Ph.D. Supervisor:
Dr. Marcella Bona

London, 2015

Declaration

The work presented in this thesis was carried out between October 2011 and December 2014. It is the result of my own studies and wherever contributions taken from other sources are involved, they are identified as such.

Cristiano Alpigiani
January 23, 2015

To my family

Contents

Introduction

The impressive progress that elementary particle physics made in the second half of the last century led to the formulation of a comprehensive theory, known as the Standard Model (SM), which correctly describes all fundamental interactions in nature, except for the gravitational one.

Indirect discoveries have always played an important role in high energy physics scenario and indirect research can be considered to all intents and purposes complementary to the direct one, since allows to test much higher energy scales than those the current colliders are able to reach. This is very important now that electroweak precision tests and measurements on Flavour Changing Neutral Currents (FCNC) processes put very stringent constraints on physics beyond the SM, requiring it to appear at scales $\mathcal{O}(10$ TeV). On the other hand, New Physics (NP) is expected already at scales $\mathcal{O}(1$ TeV) in order to offer a natural explanation to the smallness of the Higgs mass. This scale is also confirmed by recent constraints on thermal dark matter [1] which show how new physics should manifest not far above the electroweak scale.

Rare B decays have always played a crucial role in shaping the flavour structure of the SM and particle physics in general. Since the first measurement of rare radiative $B \to K^*\gamma$ decays by the CLEO Collaboration [2] this area of particle physics has received much experimental and theoretical attention. In particular, FCNC B decays, involving the b-quark transition $b \to (s,d) + \gamma$ and $b \to (s,d) + \ell^+\ell^-(\ell = e,\mu,\tau,\nu)$, provided crucial tests for the SM at the quantum level since they proceed through loop or box diagrams, and they are highly suppressed in the SM (also by helicity). Hence, these rare B decays are characterised by their high sensitivity to NP.

The $B_s^0 \to \mu^+\mu^-$ channel is the most direct example of the $b \to s\,\ell\ell$ transitions. The SM predicted branching ratio [3] can be enhanced by coupling to non-SM heavy particles, such as those predicted by the Minimal Supersymmetric Standard Model (MSSM) and other extensions. Updated measurements on the $B_s^0 \to \mu^+\mu^-$ branching ratio have been presented by ATLAS [4], LHCb [5] and CMS [6] collaborations.

In this thesis I will report all the studies I performed within the rare B decays ATLAS group, measuring the branching ratio of the $B_s^0 \to \mu^+\mu^-$ channel on data collected during LHC Run 1.

The first chapter provides a general introduction to the SM, focusing in particular on the flavour sector and the possible new physics scenarios.

Chapter 2 briefly introduces the LHC collider and the ATLAS detector, detailing

the muon and trigger systems, fundamental for the rare B decays measurements.

In chapters 3 and 4, I will summarise the work done, during my presence at CERN, on the ATLAS semiconductor strip detector, monitoring the Lorentz angle during Run 1 and measuring the backplane resistance of the silicon modules installed in the ATLAS cavern.

In chapter 5, I will review the strategy adopted to measure the $B_s^0 \to \mu^+\mu^-$ branching ratio, reporting all the studies I performed on the combinatorial background, and the results obtained on 4.9 fb^{-1} of data collected in 2011.

Chapters 6 and 7 detail respectively the additional studies I performed on the 2011 datasets and all the tests I made in preparation for the analysis on 20 fb^{-1} of data collected in 2012. I will show the studies on the discriminating variables for the rejection of the background, the tests on the multivariate analysis and on the possible strategies for the invariant mass fit for the extraction of the signal yield. All these studies proved to be fundamental for the 2012 measurement detailed in chapter 8.

Chapter 1

The Standard Model and Flavour Physics

The notion of an "Elementary Particle" is not so well-defined in High Energy Physics. It evolves with time following the progress in the experimental techniques which, by constantly increasing the resolution power of our observations, have shown that systems which were believed to be "elementary", are in fact composite of smaller constituents. In the last century, we went through the chain molecules → atoms → electrons + nuclei → electrons + protons + neutrons → electrons + quarks → ??? There is no reason to believe that there is an end in this series and, even less, that this end has already been reached.

J. Iliopoulos
ESHEP 2012, Anjou, France.

Our current knowledge in terms of particles is summarised in figure 1.1. All particles and their interactions are described as excitations of fermionic fields (spin-1/2 particles) interacting with different sets of gauge fields (spin-1 particles). The spin-1/2 particles can be grouped in three different *families*, each containing two quarks and two leptons that are respectively charged and neutral under strong interactions. The four fermions contained in each family have different combinations of electromagnetic, weak and strong charges that define uniquely their interactions (the gravitational force is not included in the SM). All "ordinary" matter that constitutes the Universe is made of protons and neutrons (which in turn are made of up and down quarks), electrons, and electron-neutrinos that are produced by fusion processes inside the stars. To the best of our knowledge, the particles present in the second and third family are an exact copy of the ones present in the first family (except for the heavier masses), and since they are unstable states, they decay into particles of the first family. The reason why in the SM we have three almost identical copies of quarks and leptons is one of the open questions in fundamental physics.

The SM is not realistic because the bare mass terms are forbidden for the electroweak

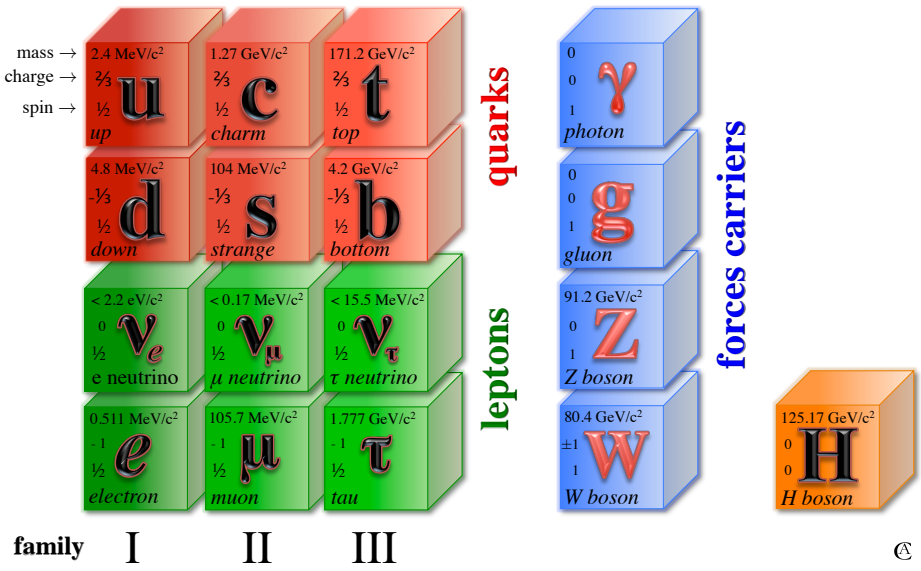


Fig. 1.1. *Fermions and bosons of the SM and their properties. Quarks are shaded red, leptons green, bosons blue and the Higgs in orange. The fermions are grouped into the three families.*

gauge bosons and for the fermions since they violate the gauge symmetry. This problem is solved by the Higgs mechanism. The masses of the vector bosons, as well as of quarks and leptons, are the result of the interaction of these fields with a new type of field, the Higgs scalar field, whose ground state spontaneously breaks the electroweak symmetry.

The SM theory has been extensively tested in the last 50 years showing an amazing agreement with measurements, recently confirmed by the discovery of the Higgs boson, made by both ATLAS [8] and CMS [9] experiments. However, there are also clear indications that this theory is not complete: the evidence for dark matter[1] and neutrino oscillations cannot be explained by the SM. Although the information currently available do not explicitly allow to unambiguously determine the range of validity of this theory, the instability of the Higgs sector under quantum corrections (fine-tuning problem) suggests that the scale where the SM is "replaced" with a more fundamental theory should be around 1 TeV, a range that can be accessible by the LHC experiments.

The description of quark and lepton masses in terms of the Higgs mechanism is unsatisfactory since the corresponding interactions are not controlled by any symmetry principle (contrary to all other known interactions), resulting in a large number of free parameters. In addition, the interaction of the quarks with the Higgs boson is responsible for the pattern of mixing of the various families of quarks under weak

[1] Recent hints for dark matter come from the measurement made by the Alpha Magnetic Spectrometer (AMS), showing an excess in the positron fraction that could point to the evidence of dark matter annihilation [10].

interactions, and the corresponding hierarchy in the decay modes.

The particular structure of the weak Higgs interactions ensure that the processes with a change of flavour mediated by Flavour Changing Neutral Currents (FCNCs) are strongly suppressed in the SM and can occur only at higher order electroweak diagrams (see section 1.5). These suppressions make the FCNCs sensitive to New Physics (NP) since new mechanisms can contribute, in addition to the SM amplitudes, generating deviations in the SM predictions for rare processes.

The next sections, based on the lectures [11] given by Gino Isidori at the 2012 European School of High Energy Physics that I attended in Anjou (France) during the first year of my PhD, will provide a general introduction to the physics of flavour, focusing in particular on rare B decays, as well as on the most plausible NP scenarios and their phenomenological implications.

1.1 “Flavour”

According to legend, the concept of “flavour physics” was coined in 1971 by Murray Gell-Mann and his student, Harald Fritzsch, at a Baskin-Robbins ice-cream store in Pasadena: *“Just as ice cream has both colour and flavour so do quarks”* [12]. In the jargon of particle physics, the term *flavour* is used to describe several copies of the same gauge representation, i.e. several fields with the same quantum numbers (the three *families* reported in figure 1.1).

The term *flavour physics* refers to the interactions that distinguish between flavours, i.e. weak and Yukawa interactions (interactions related to unbroken symmetries that are mediated by massless gauge bosons do not distinguish the flavours and thus are not part of the flavour physics).

The *flavour parameters* are parameters that carry flavour indices, that correspond to the nine masses of the charged fermions and the four mixing parameters[2] (describing the interactions between quark-antiquark pairs and $W^{\pm}$ bosons).

Flavour universal refers to the interactions with couplings (or flavour parameters) that are proportional to the unit matrix in flavour space (strong and electromagnetic interactions).

Flavour diagonal refers to the interactions with couplings (or flavour parameters) that are diagonal, but not necessary universal (Yukawa interactions of the Higgs in the mass basis).

With *flavour changing* we indicate the processes where the difference between the number of particles and the number of antiparticles of the same flavour is different in the initial and final state. In this context, the processes in which both up-type and down-type flavours, and/or both charged lepton and neutrino flavours are involved are called *flavour-changing charged-currents*, that are mediated by the W bosons and occur at tree level. The *flavour-changing neutral-currents* processes involve either up-type or down-type (but not both), and/or either charged lepton or neutrino flavours (but

[2] If we consider the Majorana mass terms for neutrinos we need to add to the list the three neutrino masses and the six mixing parameters (interactions of $W^{\pm}$ with lepton-antilepton pairs).

not both), they do not occur at tree level (within the SM) and they are often highly suppressed.

1.2 The SM Lagrangian and the Flavour Sector

A generic model of elementary particles and their interactions is defined by essentially two ingredients. The first are the symmetries of the Lagrangian and the structure of spontaneous symmetry breaking, the second are the representations of fermions and scalars.

The SM Lagrangian can be divided in three parts, the gauge sector, the Yukawa sector and the Higgs (or symmetry breaking) sector

$$\mathcal{L}_{SM} = \mathcal{L}_{\text{gauge}} + \mathcal{L}_{\text{Yukawa}} + \mathcal{L}_{\text{Higgs}} \quad . \tag{1.1}$$

The gauge sector is extremely simple and highly symmetric, it is specified by the fermions content and by the local symmetry $G = SU(3)_C \times SU(2)_L \times U(1)_Y$.

The $SU(3)$ (QCD) factor has a gauge coupling g_s and 8 gauge bosons (gluons) G^a, $a = 1, \ldots, 8$. It is non-chiral and acts on the colour indices of the L- and R-chiral quarks $q_{r\alpha}$, where $\alpha = 1, 2, 3$ refers to the colour and r to the flavour. The bare masses are not allowed in the SM and should be generated by the Higgs mechanism, while the gluons remain massless since the QCD itself is not spontaneously broken.

In contrast to QCD, the electroweak factor $SU(2)_L \times U(1)_Y$ is chiral. The group $SU(2)$ acts only on flavour indices of the L-chiral fermions, has a gauge coupling g, gauge bosons W^b ($b = 1, 2, 3$), and also includes a neutral boson W^0 associated with a fermion phase symmetry. The abelian group $U(1)$, acting on both L and R fermions (but with different charges), has a gauge coupling g' and a gauge boson B.

The gauge sector of the SM Lagrangian is

$$\begin{aligned} \mathcal{L}_{\text{gauge}} &= -\frac{1}{4} G^a_{\mu\nu} G_a^{\mu\nu} - \frac{1}{4} W^b_{\mu\nu} W_b^{\mu\nu} - \frac{1}{4} B_{\mu\nu} B^{\mu\nu} \\ &\quad + \sum_{i=1,\ldots,3} \sum_{\psi = Q^i_L, \ldots, E^i_R} i\, \bar{\psi} \not{D} \psi \end{aligned} \tag{1.2}$$

where the field strength tensors for $SU(3)$, $SU(2)$ and $U(1)$, and the covariant derivative are respectively

$$\begin{aligned} G^a_{\mu\nu} &= \partial_\mu G^a_\nu - \partial_\nu G^a_\mu + g_s\, f^{abc}\, G^b_\mu G^c_\nu \\ W^a_{\mu\nu} &= \partial_\mu W^a_\nu - \partial_\nu W^a_\mu + g\, \epsilon^{abc}\, W^b_\mu W^c_\nu \qquad (1.3) \\ B_{\mu\nu} &= \partial_\mu B_\nu - \partial_\nu B_\mu \\ D_\mu &= \partial_\mu - i g_s\, G^a_\mu\, L_a - i g\, W^b_\mu\, T_b - i g'\, B_\mu\, Y \qquad (1.4) \end{aligned}$$

The terms G^a_μ are the eight gluon fields, W^b_μ the three weak interaction bosons and B_μ the single hypercharge boson. L_a are the 3×3 Gell-Mann matrices ($\frac{1}{2}\lambda_a$ for triplets, 0 for singlets) generators of $SU(3)_C$, T_b the 2×2 Pauli matrices ($\frac{1}{2}\sigma_b$ for doublet, 0 for

singlets) generators of $SU(2)_L$, and the Y are the $U(1)_Y$ hypercharges[3] (the abelian $U(1)$ gauge boson has no self interactions). The fermion content, flavour universal and CP-conserving, consists of five fields with different quantum numbers under the gauge group[4]

$$Q_L^i(3,2)_{+1/6}\ ,\ U_R^i(3,1)_{+2/3}\ ,\ D_R^i(3,1)_{-1/3}\ ,\ L_L^i(1,2)_{-1/2}\ ,\ E_R^i(1,1)_{-1} \tag{1.5}$$

each of them appearing in three different flavours ($i = 1,2,3$). Q_L^i, U_R^i and D_R^i are the quark fields while L_L^i and E_R^i are the lepton fields. The Higgs part of the Lagrangian

$$\mathcal{L}_{\text{Higgs}} = \mu^2\,\phi^\dagger\phi - \lambda\,(\phi^\dagger\phi)^2 \tag{1.6}$$

(where ϕ is the Higgs scalar doublet, λ is the Higgs quartic coupling and μ is the mass of the field ϕ)[5] describes the scalar self-interactions, and for the SM scalar sector, where there is a single doublet, it is also CP-conserving. The local and the global symmetries are both broken by the introduction of a $SU(2)_L$ Higgs scalar doublet. While the local symmetry is *spontaneously broken* by the vacuum expectation value of the Higgs field, $\langle\phi\rangle = v = (2\sqrt{2}\,G_F)^{-1/2} \sim 174$ GeV, the global flavour symmetry is *explicitly broken* by the Yukawa interaction of ϕ with the fermionic fields:

$$-\mathcal{L}_{\text{Yukawa}} = Y_d^{ij}\,\overline{Q}_L^i\,\phi\,D_R^j + Y_u^{ij}\,\overline{Q}_L^i\,\tilde{\phi}\,U_R^j + Y_e^{ij}\,\overline{L}_L^i\,\phi\,E_R^j + \text{h.c.} \tag{1.7}$$

where $\tilde{\phi} = i\sigma_2\,\phi^\dagger$ and $Y_{u,d,e}^{ij}$ are the Yukawa matrices ($i,j = 1,2,3$). This part of the SM Lagrangian is, in general, flavour dependent (since $Y_f \not\propto \mathbf{1}$) and CP-violating.

1.2.1 Global Symmetries

In the absence of the three Yukawa matrices Y_d, Y_u and Y_e, the SM has a large $U(3)^5$ global flavour symmetry

$$G_{\text{global}}(Y_{u,d,e} = 0) = U(1)^5 \times \mathcal{G}_q \times \mathcal{G}_\ell\ \ , \tag{1.8}$$

where

$$\begin{aligned} \mathcal{G}_q &= SU(3)_{Q_L} \times SU(3)_{U_R} \times SU(3)_{D_R} \\ \mathcal{G}_\ell &= SU(3)_{L_L} \times SU(3)_{E_R} \\ U(1)^5 &= U(1)_B \times U(1)_L \times U(1)_Y \times U(1)_{PQ} \times U(1)_{E_R}\ \ . \end{aligned} \tag{1.9}$$

Three of the five $U(1)$ charges can be identified with the baryon number B, the lepton number L (both not broken by $\mathcal{L}_{\text{Yukawa}}$), and the hypercharge Y, which is gauged and broken only spontaneously by $\langle\phi\rangle \neq 0$. The two other $U(1)$ subgroups can be

[3] For a particle with charge Q and weak isospin T_3 the weak hypercharge Y is equal to $Q - T_3$.

[4] In this notation, for the generic field $\psi(A,B)_Y$, A and B are the representation under the $SU(3)_C$ and $SU(2)_L$ groups, and Y is the $U(1)_Y$ weak hypercharge.

[5] The mass of the Higgs boson is $m_H = \sqrt{-2\mu^2} = \sqrt{2\lambda v}$ where $v = -\mu^2/\lambda$ is the Higgs vacuum expectation value.

identified with the symmetry PQ (Peccei-Quinn [13]) whereby the Higgs and D_R , E_R fields have opposite charges, and with a global rotation of E_R only. This symmetry corresponds to the independent unitary rotations in flavour space of the five fermions fields reported in equation 1.5. The two non-Abelian groups $\mathcal{G}_q$ and $\mathcal{G}_\ell$ control the flavour-changing dynamics and flavour non-universality, and are explicitly broken by $Y_{d,u,e}$ not being proportional to the identity matrix.

To diagonalise each Yukawa coupling, we need two independent unitary matrices, $V_L\, Y\, V_R^\dagger = \mathrm{diag}(y_1, y_2, y_3)$. This is trivial in the lepton sector where the invariance of $\mathcal{L}_{\text{gauge}}$ under $\mathcal{G}_\ell$ allows to choose the two matrices without breaking the gauge invariance. In the quark sector, we can freely choose only three of the four unitary matrices to diagonalise Y_d and Y_u. If we choose the basis in which Y_d is diagonal (eliminating the right-handed diagonalisation matrix of Y_u) we have[6]

$$Y_d = \lambda_d \qquad Y_u = V^\dagger \lambda_u \quad , \tag{1.10}$$

where

$$\lambda_d = \mathrm{diag}(y_d, y_s, y_b) \ , \ \ \lambda_u = \mathrm{diag}(y_u, y_c, y_t) \ , \ \ y_q = \frac{m_q}{v} \tag{1.11}$$

are defined in terms of the quark masses m_q and the vacuum expectation value v. The diagonalisation from the left of both Y_d and Y_u is not allowed by flavour symmetry, thus in both cases we are left with a non-trivial unitary matrix, V, that is the Cabibbo-Kobayashi-Maskawa (CKM) mixing matrix (section 1.3).

A unitary 3×3 [N×N] complex matrix depends on three [N(N-1)/2] real rotational angles and six [N(N+1)/2] complex phases. Considering the quark basis in 1.10, the residual invariance under the flavour group allows to eliminate five of the six complex phases in V that correspond to the relative phases on the various quark fields. As a result, the physical parameters in V are four (three real angles and one complex CP phase). Therefore, the full set of parameters that control the breaking of the quark flavour symmetry is composed by the six quark masses in $\lambda_{u,d}$ and the four parameters in V.

Since V is not diagonal, the $W^\pm$ gauge bosons couple to quark mass eigenstates of different generations. Within the SM, this is the only source of flavour-changing quark interactions[7]. There is no mixing in the Z coupling, that means that there are no flavour-changing neutral currents at tree level.

1.3 The CKM Matrix and the Unitarity Triangle

As we discussed in the previous section, the quark flavour physics in the SM is characterised by a flavour symmetry of $\mathcal{G}_q$ that is broken by the Yukawa coupling Y_u and Y_d, and the CKM matrix arises by the misalignment of Y_u and Y_d in flavour space.

Many parameterisations of the CKM matrix have been proposed in literature, but we will analyse only two of them: the standard parameterisation and the Wolfenstein parameterisation [14].

[6] The same thinking can be applied to the basis where $Y_d = V\lambda_d$ and $Y_u = \lambda_u$.

[7] Since the neutrino is considered massless, there is no mixing in the lepton sector.

The standard parameterisation is

$$\begin{aligned} V &= \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} \\ &= \begin{pmatrix} c_{12}c_{13} & s_{12}c_{13} & s_{13}\,e^{-i\delta} \\ -s_{12}c_{23} - c_{12}s_{23}s_{13}\,e^{i\delta} & c_{12}c_{23} - s_{12}s_{23}s_{13}\,e^{i\delta} & s_{23}c_{13} \\ s_{12}s_{23} - c_{12}c_{23}s_{13}\,e^{i\delta} & -s_{23}c_{12} - s_{12}c_{23}s_{13}\,e^{i\delta} & c_{23}c_{13} \end{pmatrix} \end{aligned} \tag{1.12}$$

with $c_{ij} = \cos\theta_{ij}$, $s_{ij} = \sin\theta_{ij}$ $(i, j = 1, 2, 3)$ where θ_{ij} are the quark mixing angles and δ is the Kobayashi-Maskawa phase, necessary for CP violation. The terms c_{ij} and s_{ij} can all be chosen to be positive while δ can vary in the range $0 \leqslant \delta \leqslant 2\pi$ (the combination of electroweak measurements constrains δ to be in the range $0 < \delta < \pi$, as shown in figure 1.4).

The Wolfenstein parameterisation is an approximate parameterisation that considers an expansion of the CKM elements in power series of the small parameter $\lambda = |V_{us}| \sim 0.22$

$$V = \begin{pmatrix} 1 - \frac{\lambda^2}{2} & \lambda & A\,\lambda^3(\rho - i\eta) \\ -\lambda & 1 - \frac{\lambda^2}{2} & A\lambda^2 \\ A\,\lambda^3(1 - \rho - i\eta) & -A\lambda^2 & 1 \end{pmatrix} + \mathcal{O}(\lambda^4) \tag{1.13}$$

where the parameters A, ρ and η are free and of order 1. This parameterisation highlights the strongly hierarchy pattern of the off-diagonal terms ($s_{13} \ll s_{23} \ll s_{12} \ll 1$), e.g. the terms $|V_{ub}|$ and $|V_{td}|$ are of order $5 \cdot 10^{-3}$, whereas the terms $|V_{cb}|$ and $|V_{ts}|$ are of order $4 \cdot 10^{-2}$. If the level of accuracy needed is higher, also terms of $\mathcal{O}(\lambda^4)$ and $\mathcal{O}(\lambda^5)$ have to be included in the calculation.

Although the Wolfenstein parameterisation is more transparent than the standard one, in order to have an accuracy at the level of the experimental results, the higher order terms in λ should be included. The simplest choice, that allows to keep the transparency of the original Wolfenstein parameterisation, is to go back to the standard one and define the parameters λ, A, ρ and η as

$$\lambda \equiv s_{12}\,, \qquad A\lambda^2 \equiv s_{23} \qquad A\lambda^3(\rho - i\eta) \equiv s_{13}\,e^{-i\delta} \tag{1.14}$$

to all orders in λ. The change of variables $\{s_{ij}, \delta\} \to \{\lambda, A, \rho, \eta\}$ in equation 1.12 allows to obtain an exact parameterisation of the CKM matrix in terms of the Wolfenstein parameters. If we expand each element up to $\mathcal{O}(\lambda^5)$ we can obtain the following matrix

$$V = \begin{pmatrix} 1 - \frac{\lambda^2}{2} - \frac{1}{8}\lambda^4 & \lambda + \mathcal{O}(\lambda^7) & A\,\lambda^3(\rho - i\eta) \\ -\lambda + \frac{1}{2}A^2\lambda^5[1 - 2(\rho + i\eta)] & 1 - \frac{1}{2}\lambda^2 - \frac{1}{8}\lambda^4(1 + 4A^2) & A\lambda^2 + \mathcal{O}(\lambda^8) \\ A\,\lambda^3(1 - \bar{\rho} - i\bar{\eta}) & -A\lambda^2 + \frac{1}{2}A\lambda^4[1 - 2(\rho + i\eta)] & 1 - \frac{1}{2}A^2\lambda^4 \end{pmatrix} \tag{1.15}$$

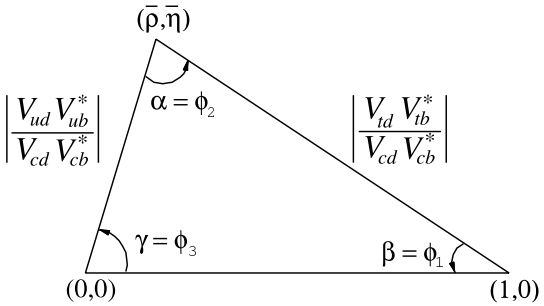


Fig. 1.2. *The CKM unitarity triangle [15].*

where

$$\bar{\rho} = \rho\left(1 - \frac{\lambda^2}{2}\right) + \mathcal{O}(\lambda^4)\,, \qquad \bar{\eta} = \eta\left(1 - \frac{\lambda^2}{2}\right) + \mathcal{O}(\lambda^4) \tag{1.16}$$

The definition of V_{ub} remains unchanged and the corrections to V_{us} and V_{cb} appear only at $\mathcal{O}(\lambda^7)$ and $\mathcal{O}(\lambda^8)$ respectively. Consequently, to an excellent accuracy we have

$$V_{us} = \lambda\,, \qquad V_{cb} = A\lambda^2 \tag{1.17}$$

$$V_{ub} = A\lambda^3(\rho - i\eta)\,, \qquad V_{td} = A\lambda^3(1 - \bar{\rho} - i\bar{\eta}) \tag{1.18}$$

Therefore, with this generalisation we do not have relevant corrections to V_{us}, V_{cb}, V_{ub} and V_{cd}, and only an elegant change in V_{td}.

The unitarity of the CKM matrix gives nine or six relations between its elements

$$1) \qquad \sum_{k=1,\cdots,3} V_{ik}^* V_{ki} = 1 \quad , \tag{1.19}$$

$$2) \qquad \sum_{k=1,\cdots,3} V_{ik}^* V_{kj\neq i} = 0 \tag{1.20}$$

whose experimental verification is useful to set constraints or, possibly, reveal the presence of NP beyond the SM. In particular, the relation obtained from equation 1.20 (for $i = 1$ and $j = 3$)

$$V_{ud}V_{ub}^* + V_{cd}V_{cb}^* + V_{td}V_{tb}^* = 0 \tag{1.21}$$

is very interesting since it involves the sum of three terms all of the same order in λ and it is usually represented as a "unitarity" triangle (see figure 1.2) in the complex plane $\bar{\rho} - \bar{\eta}$

$$\frac{V_{ud}V_{ub}^*}{V_{cd}V_{cb}^*} + \frac{V_{td}V_{tb}^*}{V_{cd}V_{cb}^*} + 1 = 0 \quad \Longleftrightarrow \quad [\bar{\rho} + i\bar{\eta}] + [(1 - \bar{\rho}) - i\bar{\eta}] + 1 = 0 \quad . \tag{1.22}$$

Since the relation 1.21 is invariant under any phase transformation in the quark fields, the corresponding triangle can be rotated in the $\bar{\rho} - \bar{\eta}$ plane, but its angles and sides remain unchanged. This implies that both the sides and the angles of the unitarity triangle are observables that can be measured in the experiments.

1.3.1 Current Status of the CKM Fits

All the measurements of rates, mixing and CP asymmetries in B decays performed at the B factories, at the Tevatron, as well as at the LHC helped to prove the validity of the Kobayashi-Maskawa (KM) mechanism and provided strong constraints on the CKM matrix parameters. As a starting point, we can assume the SM and check its overall consistency, but the richness of data collected allows to go further and test the contributions of potential NP to the processes we are studying.

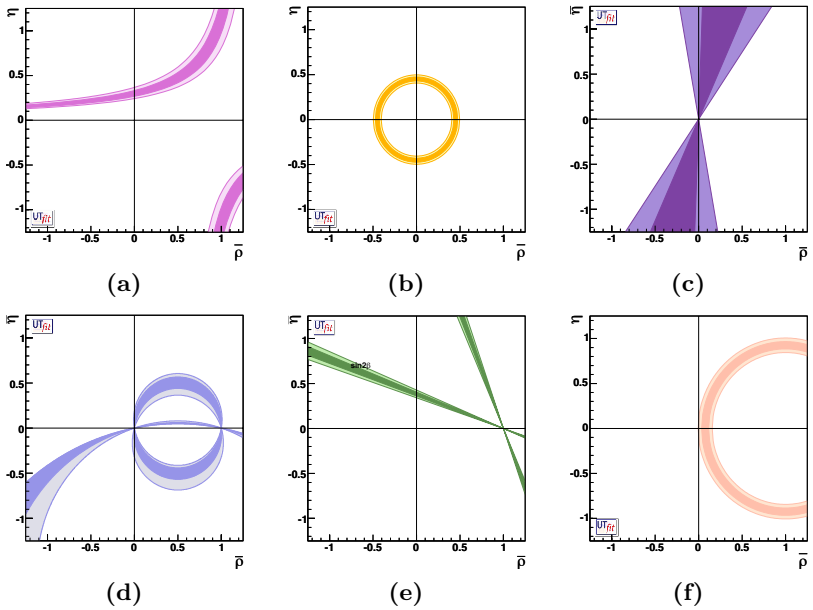


Fig. 1.3. *Most sensitive observables used to constrain the SM [16]. From left to right, from top to bottom: CP violating parameter of the kaon system ϵ_K (a), rates of inclusive and exclusive charmless semi-leptonic B decays (b), rates of various $B \to D^{(*)}K$ decays (c), rates of various $B \to \pi\pi$, $\rho\pi$, $\rho\rho$ decays (d), time-dependent CP asymmetry in the $B \to J/\psi\, K_S$ decays (e), ratio between the mass splittings in the neutral B^0 and B_s system (f).*

The values of λ and A (i.e. $|V_{us}|$ and $|V_{cb}|$) are measured with a very good accuracy in the decays $K \to \pi\,\ell\,\nu$ and $B \to X_c\,\ell\,\nu$. The latest measurements from the UTfit Collaboration [16] are

$$\lambda = 0.2255 \pm 0.0005\,, \qquad A = 0.820 \pm 0.012\,. \tag{1.23}$$

Using these results, all the other constraints on the elements of the CKM matrix can be expressed as constraint on $\bar{\rho}$ and $\bar{\eta}$ (i.e. constraints on the CKM unitarity triangle). Figure 1.3 reports the most sensitive observables that can be used to constrain the SM:

(a) The mixing CP violating parameter of the kaon system, ϵ_K, that determines an hyperbola in the $\bar{\rho} - \bar{\eta}$ plane.

(b) The rates of inclusive and exclusive charmless semi-leptonic B decays that determines $|V_{ub}|^2 \propto \bar{\rho}^{\,2} + \bar{\eta}^{\,2}$.

(c) The rates of various $B \to D^{(*)}K$ decays that constrain the angle γ, where $e^{i\gamma} = (\bar{\rho} + i\bar{\eta})/(\bar{\rho}^2 + \bar{\eta}^2)$.

(d) The rates of various $B \to \pi\pi$, $\rho\pi$, $\rho\rho$ decays that constrain the phase $\alpha = \pi - \beta - \gamma$.

(e) The time-dependent CP asymmetry in the $B \to J/\psi\, K_S$ decays $(\mathcal{A}^{CP}_{KJ/\psi})$[8] that constrain the $\sin 2\beta = 2\bar{\eta}\,(1-\bar{\rho})/((1-\bar{\rho})^2 + \bar{\eta}^2)$.

(f) The ratio between the mass splittings in the neutral B^0 and B_s system that depends on $|V_{td}/V_{ts}|^2 \propto [(1-\bar{\rho})^2 + \bar{\eta}^2]$.

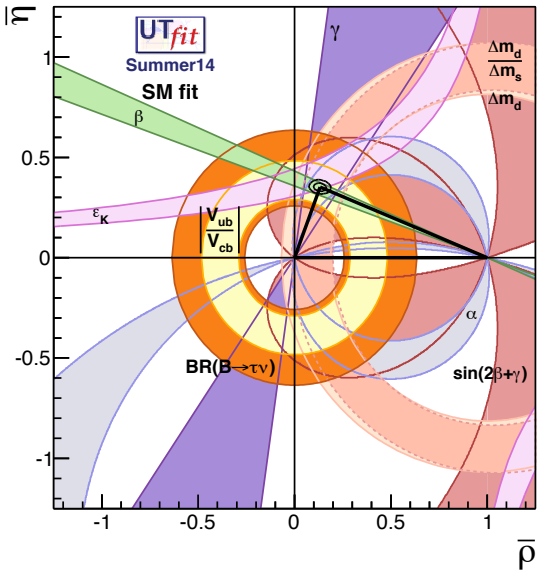


Fig. 1.4. *Determination of $\bar{\rho}$ and $\bar{\eta}$ from constraints on $|V_{ub}|/|V_{cb}|$, Δm_d, Δm_s, ϵ_K , β, γ, and α. 68 % and 95 % total probability contours are shown, together with 95 % probability regions from the individual constraints [16].*

The result of all these constraints, obtained by the UTfit Collaboration [16][9], is reported in figure 1.4 that shows how they are all consistent with a unique value in the $\bar{\rho} - \bar{\eta}$ plane

$$\bar{\rho} = 0.132 \pm 0.023\,, \qquad \bar{\eta} = 0.351 \pm 0.013 \quad . \tag{1.24}$$

This impressive consistency of the different constraints on the CKM unitarity triangle shows how well the SM is able to describe the flavour-changing phenomena, leaving only little room for non-SM contributions.

[8] The CP asymmetry depends on the phase of the $B_d - \overline{B}_d$ mixing amplitude relative to the decay amplitude.

[9] Similar results can be obtained by another collaboration, CKMfitter [17].

1.4 The New Physics Flavour Puzzle

Despite the impressive phenomenological success of the SM in describing the flavour and electroweak physics, it is clear how it cannot be considered a complete theory of Nature, but rather an approximation in the low-energy limit of a more complete theory. The SM does not include gravity and therefore can not be valid at energy scale above $M_{\text{Plank}} \sim 10^{19}$ GeV. On the other hand, the evidence of the presence of dark matter in the Universe and the fine-tuning problem of the Higgs mass suggest that the energy scale Λ_{NP} at which the SM is replaced by a more fundamental theory should be much lower than the above one, $\Lambda_{\text{NP}} \lesssim 1$ TeV.

If we assume that the new degrees of freedom, which complete the theory, are heavier than the SM particles, we can integrate them out and describe the NP model using an *effective theory.* In quantum field theory, an effective theory is a description of physics at a given scale in terms of the degrees of freedom that can actually appear as physical states at that energy. In these models the SM Lagrangian becomes the renormalisable part of a more general Lagrangian which includes a series[10] of non-renormalisable operators with dimension $d > 4$. These new heavy particles, which do not appear explicitly in the theory, but whose effects are described by the new operators, are expressed in terms of SM fields and suppressed as $1/\Lambda_{\text{NP}}$. The problem of integrating out these degrees of freedom is that we do not know the nature of them and thus we are not able to evaluate the effective couplings of the higher-dimensional operators. Therefore, with an effective theory, we can analyse all possible extensions of the SM in terms of a limited number of coefficients of the higher-dimensional operators, but we cannot define the correlations of NP effects at low and high energies.

Considering a single elementary Higgs field responsible for $SU(2)_L \times U(1)_Y \to U(1)_Q$ symmetry breaking, the most general Lagrangian of the SM as an effective theory can be written as

$$\mathcal{L}_{\text{eff}} = \mathcal{L}_{\text{gauge}} + \mathcal{L}_{\text{Yukawa}} + \mathcal{L}_{\text{Higgs}} + \Delta\mathcal{L}_{d>4} \quad , \tag{1.25}$$

where

$$\Delta\mathcal{L}_{d>4} = \sum_{d>4} \sum_{i=1}^{N_d} \frac{C_i}{\Lambda^{d-4}} Q_i^{(d)} \,(\text{SM fields}) \tag{1.26}$$

is a combination of operators ($Q_i^{(d)}$) that contain only the SM fields and are invariant under the SM gauge group. The terms C_i are effective coupling while Λ is the energy scale. Since we expect NP at the TeV scale, if the underlying theory is natural, i.e. there is no fine-tuning in the coupling constant, we expect $C_i = \mathcal{O}(1)$ for all the operators that are not suppressed or forbidden by symmetry arguments.

One of the strategies to obtain clues of the value of Λ_{NP} is to constrain (or even better find evidence) the effective non-renormalisable interactions that encode the presence of new degrees of freedom at higher energies. The non-renormalisable operators

[10] There are 59 dimension-six independent operators (5 additional operators can be considered assuming no baryon number conservation in the four-fermion sector). There is only one SM dimension-five operator (the well-known Weinberg operator [18]) which leads to neutrino masses and that violates the lepton number.

should naturally induce large effects in FCNC processes ($\Delta F = 1$ and $\Delta F = 2$) that have not been seen up to now. This implies severe bounds on the effective scale of the dimension-six FCNC operators. For example, the good agreement between the SM and the measurements of $K^0 - \overline{K}^0$ mixing leads to a bound higher than 10^3 TeV, well above the scale suggested by the Higgs sector.

The contradiction between the above determination of Λ_{NP} and the expectation derived from natural solution of the hierarchy problem is what is called *flavour problem*. If we believe that new physics should be present at the TeV scale, the theory should have a highly non-generic flavour structure (see section 1.7). Since the SM does not have an exact flavour symmetry (being broken by the Yukawa interactions), we need to constrain its form using the experimental results on FCNC.

The good agreement between the SM and the experimental measurements does not imply that the flavour physics is not interesting anymore. There are many other measurements that can reveal additional key features of potential physics beyond the SM, and an increased experimental precision will be extremely useful for this purpose.

1.5 FCNC Processes

The measurements of mixing-induced CP violation and tree-level allowed semi-leptonic transitions provide consistency checks of the SM and a precise determination of the parameters of the CKM matrix. The next goal is try to understand if there is still room for new sources of flavour symmetry breaking close to the electroweak scale. The decay amplitude dominated by the electroweak dynamics (that translates into an enhanced sensitivity to non-standard contributions), and the small theoretical error, due to a good control of both perturbative and non-perturbative corrections, make the rare B decays, mediated by FCNC, a very powerful tool to test the presence of NP. In the next sections we will give a general introduction to the physics behind these processes (for more details refer to [14] and [19]).

Flavour-changing neutral currents are similar to the electric dipole moments and CP violation: they are strongly suppressed in the SM, but they can be much larger in most extensions.

Within the SM the couplings of the Z, γ and G to fermions are flavour diagonal at tree level because of the GIM mechanism (Glashow-Iliopoulos-Maiani, 1970 [20]), i.e. all fermions that have the same charge, colour, chirality, and thus are able to mix with each other, are assigned to the same kind of $SU(2) \times U(1)$ representation. However, with the help of the $W^{\pm}$-vertex it is possible to construct one-loop and higher order diagrams which mediate the FCNC processes. At one-loop level they can be described by a set of basic triple and quartic effective vertices that are called *penguin* and *box* diagrams (see figure 1.5).

1.5.1 Effective Lagrangians

The decays of B mesons involve two widely separated scales: the electroweak scale (characterised by the W boson mass) that is responsible for the flavour-changing

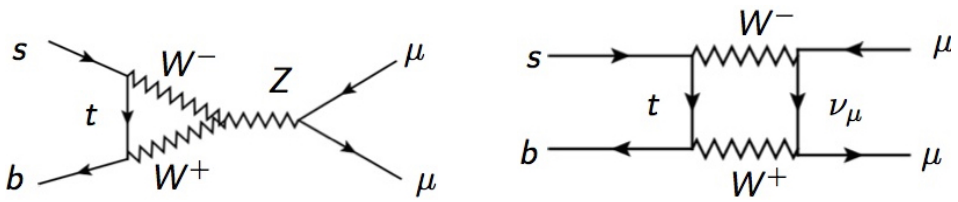


Fig. 1.5. *Example of penguin (left) and box (right) diagrams.*

transition at the quark level, and the scale of the strong interactions Λ_{QCD} that is related to the hadron formation. In this scenario, due to the wide separation between the two scales, the calculation of the decay amplitude starting from the full SM Lagrangian becomes quite complicated. This can be simplified using the Operator Product Expansion (OPE) [21, 22] that allows to describe these processes by effective weak Lagrangians where the SM fields (W, Z bosons as well as quark top) are eliminated as dynamical degrees of freedom from the theory [23, 24]. The low-energy Lagrangian contains six-dimension (and higher) local operators, expressed in terms of SM fermions, photon and gluon fields that are suppressed by power series of $1/m_W$. The OPE realises the scale separation between short-distance (high energy) and long-distance (low-energy) physics, and the scale μ, at which the local operators are renormalised, set the threshold between the two regimes.

The effective Lagrangian for a process $\Delta F = 1$ is

$$\mathcal{L}_{\Delta F=1} = -4\frac{G_F}{\sqrt{2}} \sum_i C_i(\mu) Q_i \quad , \tag{1.27}$$

where the sum is over a complete basis of operators Q_i (see equation 1.30). The terms $C_i(\mu)$ are the effective coupling, known as Wilson coefficients [25], that in general depend on the renormalisation scale and contain short-distance strong-interaction effects that can be computed using renormalisation-group techniques. When we evaluate the matrix elements of the effective Lagrangian

$$\mathcal{M}(i \to f) = -4\,\frac{G_F}{\sqrt{2}} \sum_i C_i(\mu) \langle f|Q_i(\mu)|i\rangle \tag{1.28}$$

the dependence on the scale μ cancels out. This is valid for any initial and final state, including partonic states at high energies.

1.5.2 Effective Operators for Rare Decays

The effective Lagrangian in the transitions where the underlying parton process is $b \to s + q\bar{q}$ can be written as

$$\mathcal{L}^{\text{non-lept}}_{b\to s} = -4\,\frac{G_F}{\sqrt{2}} \left(\sum_{q=u,c} \lambda^s_q \sum_{i=1,2} C_i(\mu) Q^q_i(\mu) - \lambda^s_t \sum_{i=3}^{10} C_i(\mu) Q_i(\mu) \right) \quad , \tag{1.29}$$

where $\lambda_q^s = V_{qb}^* V_{qs}$, and the operator basis is

$$\begin{aligned}
Q_1^q &= \bar{b}_L^\alpha \gamma^\mu q_L^\alpha \, \bar{q}_L^\beta \gamma_\mu s_L^\beta \,, & Q_2^q &= \bar{b}_L^\alpha \gamma^\mu q_L^\beta \, \bar{q}_L^\beta \gamma_\mu s_L^\alpha \,, \\
Q_3 &= \bar{b}_L^\alpha \gamma^\mu s_L^\alpha \, \textstyle\sum_q \bar{q}_L^\beta \gamma_\mu q_L^\beta \,, & Q_4 &= \bar{b}_L^\alpha \gamma^\mu s_L^\beta \, \textstyle\sum_q \bar{q}_L^\beta \gamma_\mu q_L^\alpha \,, \\
Q_5 &= \bar{b}_L^\alpha \gamma^\mu s_L^\alpha \, \textstyle\sum_q \bar{q}_R^\beta \gamma_\mu q_R^\beta \,, & Q_6 &= \bar{b}_L^\alpha \gamma^\mu s_L^\beta \, \textstyle\sum_q \bar{q}_R^\beta \gamma_\mu q_R^\alpha \,, \\
Q_7 &= \tfrac{3}{2} \bar{b}_L^\alpha \gamma^\mu s_L^\alpha \, \textstyle\sum_q e_q \bar{q}_R^\beta \gamma_\mu q_R^\beta \,, & Q_8 &= \tfrac{3}{2} \bar{b}_L^\alpha \gamma^\mu s_L^\beta \, \textstyle\sum_q e_q \bar{q}_R^\beta \gamma_\mu q_R^\alpha \,, \\
Q_9 &= \tfrac{3}{2} \bar{b}_L^\alpha \gamma^\mu s_L^\alpha \, \textstyle\sum_q e_q \bar{q}_L^\beta \gamma_\mu q_L^\beta \,, & Q_{10} &= \tfrac{3}{2} \bar{b}_L^\alpha \gamma^\mu s_L^\beta \, \textstyle\sum_q e_q \bar{q}_L^\beta \gamma_\mu q_L^\alpha \,,
\end{aligned} \tag{1.30}$$

with $\{\alpha, \beta\}$ and e_q being respectively the colour indices and the electric charge of the quark $q\,(= u, c)$, while γ^μ are the standard Dirac matrices. Only the two operators Q_1^c and Q_1^u are generated at tree level by the exchange of the W boson. The operators Q_7 - Q_{10} are the more interesting ones since the initial conditions of their coefficients are related to the electroweak penguin and box diagrams. Unfortunately, the contributions from long-distance physics dilute the interesting short-distance information, and thus NP effects could be hard to measure here.

If we consider the $b \to s$ transitions with a photon or a lepton pair $(l^+ l^-)$ in the final state, we need to add four additional dimension-six operators to the basis reported in 1.30,

$$\mathcal{L}_{b\to s}^{\text{rare}} = \mathcal{L}_{b\to s}^{\text{non-lept}} + 4 \frac{G_F}{\sqrt{2}} \lambda_t^s \left(C_{7\gamma} Q_{7\gamma} + C_{8g} Q_{8g} + C_{9V} Q_{9V} + C_{10A} Q_{10A} \right) \,, \tag{1.31}$$

where

$$\begin{aligned}
Q_{7\gamma} &= \frac{e}{16\,\pi^2} \, m_b \, \bar{b}_R^\alpha \, \sigma^{\mu\nu} F_{\mu\nu} \, s_L^\alpha \,, & Q_{8g} &= \frac{g_s}{16\,\pi^2} \, m_b \, \bar{b}_R^\alpha \, \sigma^{\mu\nu} G_{\mu\nu}^a \, L_a s_L^\alpha \,, \\
Q_{9V} &= \frac{1}{2} \, \bar{b}_L^\alpha \gamma^\mu s_L^\alpha \, \bar{l} \, \gamma_\mu \, l \,, & Q_{10A} &= \frac{1}{2} \, \bar{b}_L^\alpha \gamma^\mu s_L^\alpha \, \bar{l} \, \gamma_\mu \gamma_5 \, l \,.
\end{aligned} \tag{1.32}$$

$F_{\mu\nu}$ and $G_{\mu\nu}^a$ are the photon and gluon field strength tensors respectively, and $\sigma^{\mu\nu} = i\,[\gamma^\mu, \gamma^\nu]/2$. In the SM, these operators are generated by one-loop penguins and box diagrams dominated by the top quark, and are particularly sensitive to NP. Among the four reported in 1.32, the most clean is the axial-current operator

$$C_{10A}^{\text{SM}}(m_W) = \frac{g^2}{8\,\pi^2} \, \frac{x_t}{8} \left[\frac{4 - x_t}{1 - x_t} + \frac{3\,x_t}{(1 - x_t)^2} \, \ln x_t \right] \quad \text{where} \quad x_t = \frac{m_t^2}{m_W^2} \tag{1.33}$$

that does not mix with any of the four-quark operators. If new physics is present at the TeV scale, it can modify this result that would lead to a change of the value of the low-energy observables sensitive to C_{10A}, like the branching ratio of $B \to \ell^+ \ell^-$ and $\mathcal{A}_{\text{FB}}(B \to K^* \ell^+ \ell^-)$.

1.6 Standard Model Prediction for the $B_{s,d} \to \mu^+ \mu^-$ BR

In general, the predictions for the FCNC decays are difficult due to non-perturbative effects that have to be kept well under control. An exception is represented by the

very clear pure leptonic decays of the B_s and B_d into two leptons, where the non-perturbative effects are encoded in the meson decay constant. Their branching ratios undergo an additional helicity suppression by $m_\ell^2/m_{B_q}^2$, where m_ℓ and m_{B_q} are the masses of the charged lepton and the B_q meson respectively. This suppression can be lifted in models such the minimal supersymmetric SM.

Going into the detail of the $B_{s,d} \to \mu^+\mu^-$ channels, the most updated experimental world average branching ratios are

$$\mathrm{BR}(B_s \to \mu^+\mu^-) = 2.8^{+0.7}_{-0.6} \times 10^{-9} \qquad \mathrm{BR}(B_d \to \mu^+\mu^-) = 3.9^{+1.6}_{-1.4} \times 10^{-10} \quad (1.34)$$

obtained by the combination of CMS and LHCb results [28]. The statistical significance from the likelihood is 6.2 σ for the $B_s \to \mu^+\mu^-$ (7.4 σ expected from SM) and 3.2 σ for the $B_d \to \mu^+\mu^-$ (0.8 σ expected from SM). The 2D likelihood scan of $\mathcal{S} = \mathrm{BR}/\mathrm{BR}^{\mathrm{SM}}$ gives

$$\mathcal{S}^{B_s}_{\mathrm{SM}} = 0.76^{+0.20}_{-0.18} \qquad \mathcal{S}^{B_d}_{\mathrm{SM}} = 3.7^{+1.6}_{-1.4} \quad (1.35)$$

that correspond to a compatibility with the SM of 1.2 σ for the B_s and 2.2 σ for the B_d (the theoretical errors have been included in the fit).

From the most updated theoretical calculation [3], evaluated including the NLO electroweak and NNLO QCD corrections, the B_s branching ratio can be written as

$$\mathrm{BR}(B_s \to \mu^+\mu^-)^{\mathrm{SM}} \times 10^9 = (3.65 \pm 0.06) \cdot R_{t\alpha} \cdot R_s = 3.65 \pm 0.23 \quad (1.36)$$

where

$$R_{t\alpha} = \left(\frac{m_t}{173.1\,[\mathrm{GeV}]}\right)^{3.06} \left(\frac{\alpha_s(m_Z)}{0.1184}\right)^{-0.18} \quad (1.37)$$

$$R_s = \left(\frac{f_{B_s}\,[\mathrm{MeV}]}{227.7}\right)^2 \cdot \left(\frac{|V_{cb}|}{0.0424}\right)^2 \cdot \left(\frac{|V_{tb}^* V_{ts}/V_{cb}|}{0.980}\right)^2 \cdot \frac{\tau_H^s\,[\mathrm{ps}]}{1.615} \quad (1.38)$$

with $\tau_H^s \equiv 1/\Gamma_H^s$ [29] where Γ_H^s denotes the heavier mass-eigenstate total width of the $B_s^0 - \overline{B}_s^0$ system. This allows to avoid considering correlations between the decay width difference and the average lifetime. The correlations between the B_s decay constant f_{B_s} and α_s have been ignored, whereas the uncertainties related to the parameters that do not occur in the quantities $R_{t\alpha}$ and R_s have been absorbed into the residual error in the middle term of equation 1.36 that is dominated by a non parametric uncertainty set to 1.5 % of the branching ratio[11] (see [3] for more details).

Along the same line, it is possible to calculate the branching ratio for the B_d channel

$$\mathrm{BR}(B_d \to \mu^+\mu^-)^{\mathrm{SM}} \times 10^{10} = (1.06 \pm 0.02) \cdot R_{t\alpha} \cdot R_d = 1.06 \pm 0.09 \quad (1.39)$$

with

$$R_d = \left(\frac{f_{B_d}\,[\mathrm{MeV}]}{190.5}\right)^2 \cdot \left(\frac{|V_{tb}^* V_{td}|}{0.0088}\right)^2 \cdot \frac{\tau_{\mathrm{av}}^d\,[\mathrm{ps}]}{1.519} \quad , \quad (1.40)$$

[11] The previous measurements had a non parametric uncertainty of around 8 %.

where $1/\Gamma_H^d \simeq 2/(\Gamma_H^d + \Gamma_L^d) \equiv \tau_{\rm av}^d$ given the tiny SM expectation for $(\Gamma_L^d - \Gamma_H^d)/(\Gamma_L^d + \Gamma_H^d) \equiv \Delta\Gamma^d/(2\Gamma_{\rm av}^d) = 0.0021$ [30].

The main parametric uncertainties in the evaluation of the B_q branching ratios come from f_{B_q} and the CKM parameters.

The increased accuracy in the theoretical estimation of the branching ratios is essential to interpret the experimental results in terms of SM or NP contributions. This will be quite important in the next years when the experimental accuracy will be able to reach the same level of the current theoretical one.

The expected theoretical fraction between B_d and B_s branching ratios can be expressed as [3, 31–34]

$$\mathcal{R} = \frac{\mathrm{BR}(B_d \to \mu^+\mu^-)^{\rm SM}}{\mathrm{BR}(B_s \to \mu^+\mu^-)^{\rm SM}} = \frac{\tau_{\rm av}^d}{\tau_H^s} \cdot \left(\frac{f_{B_d}}{f_{B_s}}\right)^2 \cdot \left|\frac{V_{td}}{V_{ts}}\right|^2 \cdot \frac{M_{B_d}\sqrt{1-\frac{4m_\mu^2}{M_{B_d}^2}}}{M_{B_s}\sqrt{1-\frac{4m_\mu^2}{M_{B_s}^2}}} = 0.0295^{+0.0028}_{-0.0025} \tag{1.41}$$

and, given the most updated value of the ratio of the hadronisation fractions [35], $f_d/f_s = 3.83 \pm 0.22$, we can now evaluate the ratio of the expected number of B_d and B_s

$$\frac{N(B_d)^{\rm SM}}{N(B_s)^{\rm SM}} = \mathcal{R} \cdot \frac{f_d}{f_s} = 0.113 \pm 0.013 \ . \tag{1.42}$$

Coming back to the experimental results, the combination of the CMS + LHCb shows a ratio between the two channels higher with respect to the SM prediction

$$\mathcal{R} = 0.14^{+0.08}_{-0.06} \ , \tag{1.43}$$

compatible with the SM at 2.3 σ (considering the theoretical uncertainty in the evaluation). Thus, the measured ratio of number of events of B_d and B_s becomes

$$\frac{N(B_d)^{\rm SM}}{N(B_s)^{\rm SM}} = 0.536 \pm 0.308 \ . \tag{1.44}$$

Despite the above deviations from the mean value, the experimental measurements are still compatible with the SM predictions. The new data collected by Run 2 will be, therefore, fundamental to improve the precision and confirm, or contradict, the presence of NP in the flavour sector.

1.7 Flavour Physics Beyond the Standard Model

Assuming that the NP respects the SM gauge symmetry, the corrections to low-energy flavour-violating amplitudes can be written as

$$\mathcal{A}(f_i \to f_j + X) = A_0 \left(\frac{C_{\rm SM}}{m_W^2} + \frac{C_{\rm NP}}{\Lambda^2}\right) \ , \tag{1.45}$$

where Λ is the usual energy scale of the new degrees of freedom. The coefficients $C_{\rm SM}$ and $C_{\rm NP}$ can include CKM factors and eventually a $\sim 1/(16\pi^2)$ suppression [11, 36] if the amplitude is loop-mediated. Since we do not know the value of $C_{\rm NP}$, the scale Λ, given the current experimental results, varies over a wide range. Equation 1.45 allows to predict how the future experiments can improve the bound of NP. Since the sensitivity on Λ goes as $N^{1/4}$, where N is the number of events used to measure the observable, we can easily understand how it is difficult to increase substantially the energy limit with indirect NP searches only. For models where $C_{\rm SM} \ll C_{\rm NP}$ (i.e. models that do not respect the symmetries and the symmetry-breaking pattern of the SM) the indirect search can probe NP scales well above the TeV.

The analysis of the bounds on $\Delta F = 2$ operators [37] shows how the current data probe very high scales. On the other hand, if we want to keep our firm belief that NP should be present not far from TeV scale (in order to stabilise the Higgs sector), then the new degrees of freedom should have a flavour structure that can explain the smallness of the effective coupling $C_{\rm NP}$ for $\Lambda = 1$ TeV.

Potential NP may affect the unitarity triangle results shown in figure 1.4. Starting from the two quantities that are largely free from any impact of non-standard model physics, i.e. $|V_{ub}/V_{cb}|$ measured from semi-leptonic decays ($b \to c\,\ell\,\bar{\nu}_\ell$ and $b \to u\,\ell\,\bar{\nu}_\ell$, dominated by SM W boson exchange) and the angle γ from $B \to D^{(*)}\,K$ decays, we can constrain, as reported in figure 1.6, the allowed values in the $\bar{\rho} - \bar{\eta}$ plane. From these bounds it is possible to set a prediction for all observables of the unitarity triangle that each model beyond the SM has to respect in order to be reasonable and compatible with the precise measurements obtained up to now.

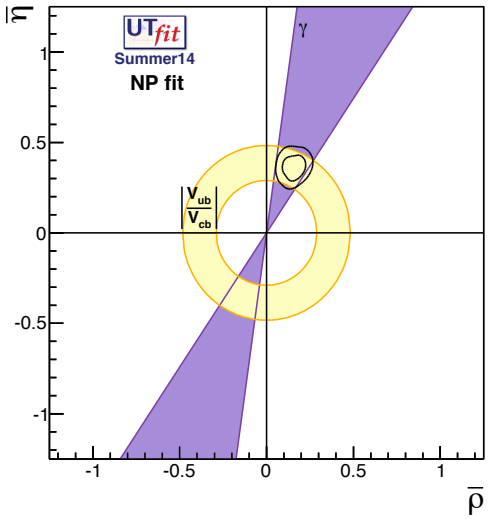


Fig. 1.6. *Constraint to NP models from γ and $|V_{ub}/V_{cb}|$ [39].*

1.7.1 $B_{s,d} \to \mu^+\mu^-$ as Probe for New Physics

As we have already seen, the $B_{s,d} \to \mu^+\mu^-$ decays are strongly helicity-suppressed in the SM and their branching ratio could be enhanced by the presence of NP in the scalar and pseudo-scalar operators. This makes these processes an excellent and precise probe to investigate the reliability of the theories beyond the SM.

Several different models describe the potential new physics scenarios. We will not analyse all of them in detail, but we will rather show how the rare B decays measurements can constrain their parameters.

Operator	Bound in Λ	Observable
$(\overline{Q}_L Y^u Y^{u\dagger} \gamma_\mu Q_L)(\overline{E}_R \gamma_\mu E_R)$	2.7 TeV	$B_s^0 \to \mu^+\mu^-$, $B \to X_s \ell^+\ell^-$
$i\,(\overline{Q}_L Y^u Y^{u\dagger} \gamma_\mu Q_L)(H_U^\dagger D_\mu H_U)$	2.3 TeV	$B_s^0 \to \mu^+\mu^-$, $B \to X_s \ell^+\ell^-$
$(\overline{Q}_L Y^u Y^{u\dagger} \gamma_\mu Q_L)(\overline{L}_L \gamma_\mu L_L)$	1.7 TeV	$B_s^0 \to \mu^+\mu^-$, $B \to X_s \ell^+\ell^-$

Table 1.1. *Bounds on the scale of new physics (at 95 % CL) for MFV operators $\Delta F = 1$ whose corresponding observable is the $B_s^0 \to \mu^+\mu^-$ branching ratio [37].*

The Minimal Flavour Violation (MFV) [38] hypothesis seems to be the most reasonable (but also most pessimist) solution to the flavour problem. The idea is that flavour- and CP-violating interactions are linked to the known structure of Yukawa coupling also beyond the SM. In this framework, the deviations from the SM in FCNC amplitudes rarely exceed the $\mathcal{O}(10\,\%)$ level (or the level of irreducible theoretical errors in most of the available observables). In this case, several of the constraints used to determine the CKM matrix (the unitarity triangle in particular) are not affected by NP, whose effects are not only negligible at tree-level, but also in a few clean observables sensitive to loop effects like the CP violation asymmetry in $B_d \to J/\psi\, K_S$. For example, the basic flavour-changing coupling in MFV implies that the weak CP violating phase of the B_d-$\overline{B}_d$ mixing, $(V_{tb}V_{tb}^*)^2$, is exactly the same of the SM. This can justify why most of the clean observables measured at the B factories are insensitive to NP effects in MFV framework. Figure 1.7 shows the CKM unitarity triangle fit within the SM compared with the *universal unitarity triangle* that do not have any new operators beyond those present in the SM and in which all flavour changing transitions are governed by the CKM matrix with no new phases beyond the CKM phase. Since only ϵ_K and Δm_{B_d} (but not the ratio $\Delta m_{B_d}/\Delta m_{B_s}$) are sensitive to NP within the MFV models the universal triangle can be determined simply removing the information related to ϵ_K and Δm_{B_d} from the full UTfit, but still considering the ratio $\Delta m_{B_d}/\Delta m_{B_s}$. Table 1.1 shows the bounds on the scale of new physics for MFV operators provided by the $B_s^0 \to \mu^+\mu^-$ measurements. The MFV prediction for the branching ratio of the $B_d \to \mu^+\mu^-$ channels is $< 1.2 \cdot 10^{-9}$.

The Minimal Supersymmetric extension of the SM (MSSM) [41] is, doubtless, the most studied extension of the SM at the TeV scale. The MSSM is minimal in the sense that the fewest number of additional new particles and interactions consistent with phenomenology are added to the model. The MSSM consists of the SM gauge

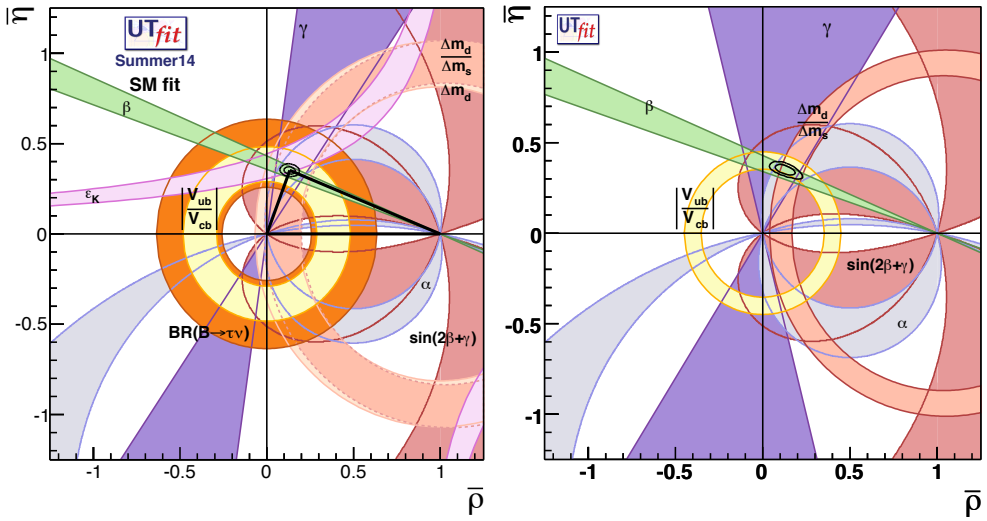


Fig. 1.7. *Comparison between the CKM unitarity triangle fit (left) and the universal unitarity triangle (right) [39].*

and fermion fields plus a scalar partner for each quark and lepton (called *squarks* and *sleptons*) and spin-1/2 partner for each gauge field (*gauginos*). The Higgs sector has two Higgs doublets with the corresponding spin-1/2 partner for each gauge field (*higgsinos*). The gauge and Yukawa interactions are completely specified in terms of the corresponding SM couplings, whereas the supersymmetry breaking is implemented by the explicit introduction of soft-supersymmetry breaking terms that can be divided into two classes: the first one contains mass terms (like sfermion or gauginos mass terms), while the second contains trilinear couplings among the scalar fields of the theory.

An example of SUSY model with MFV is the constrained supersymmetric extension of the SM (CMSSM) [42–44], known also as mSUGRA. In addition to the SM couplings, this model assumes that, at the scale of Grand Unification ($M_{GUT} \sim 10^{16}$ GeV), there are only three independent soft breaking terms: the universal gaugino mass $m_{1/2}$, the universal trilinear term A_0 and the universal sfermion mass m_0. The model has two additional free parameters in the Higgs sector (the higgsino mass term μ and B_0 which control the vacuum expectation values of the two Higgs fields). By imposing the correct masses of the W and Z bosons, we can eliminate one of the two additional free parameters and the remaining is usually chosen to be $\tan\beta = v_u/v_d$ which is the ratio of the two Higgs vacuum expectation values. In this way we can describe the model in terms of only the three high-energy parameters ($m_{1/2}$, m_0, A_0) and the low-energy parameter $\tan\beta$. Due to the very low number of parameters, the CMSSM is very predictive and can be constrained by the precise measurements in flavour physics.

In this framework, the $B_s^0 \to \mu^+\mu^-$ channel represents a very powerful tool to test the large values of $\tan\beta$ as shown in figure 1.8, where its branching ratio is

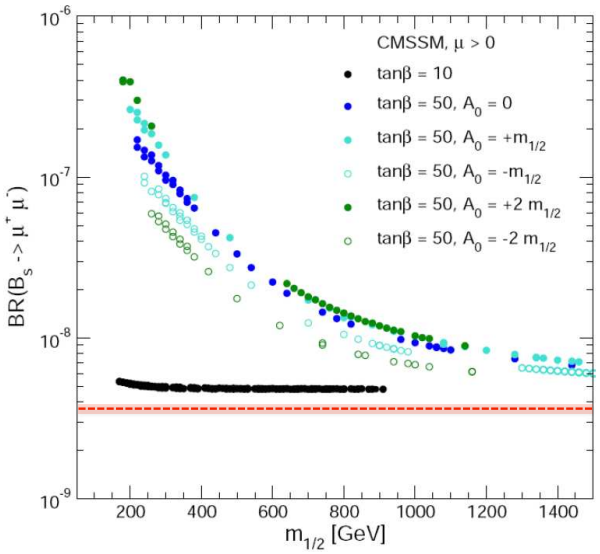


Fig. 1.8. *Branching ratio of $B_s^0 \to \mu^+\mu^-$ as a function of the gaugino mass ($m_{1/2}$) in the CMSSM model. Different predictions for the values of $\tan\beta$ and A_0 are reported (plot from [40]). The dashed red line and the shaded band show respectively the value and the error of the branching ratio obtained by the combination of CMS and LHCb measurements [28].*

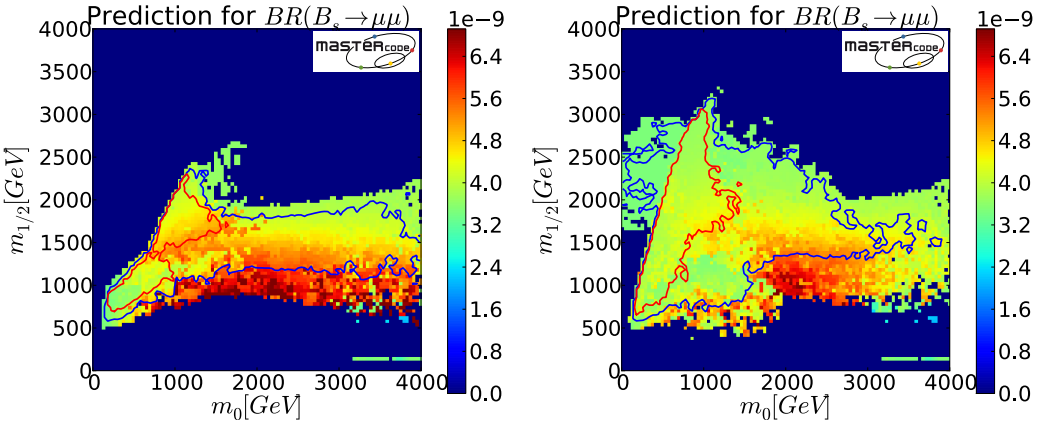


Fig. 1.9. *Prediction of the $B_s^0 \to \mu^+\mu^-$ branching ratio in the $m_0 - m_{1/2}$ plane for CMSSM (left) and NUHM1 (right) models. The red (blue) contour shows the 68 % (95 %) CL exclusion obtained including the results from LHCb, CMS, ATLAS and CDF available in 2012 [45].*

reported as a function of the gaugino mass for different predictions of $\tan\beta$ and A_0 (the horizontal dashed red line is the value obtained by the combination of LHCb and CMS measurements [28]).

In more detail, figures 1.9 and 1.10 show respectively the $B_s^0 \to \mu^+\mu^-$ branching ratio prediction in the plane $m_0 - m_{1/2}$ and $M_A - \tan\beta$ (M_A is the mass of the pseudo-scalar Higgs) for the CMSSM and NUHM1. The NUHM1 model is a generalisation of the CMSSM that includes common but non-universal soft supersymmetry-breaking

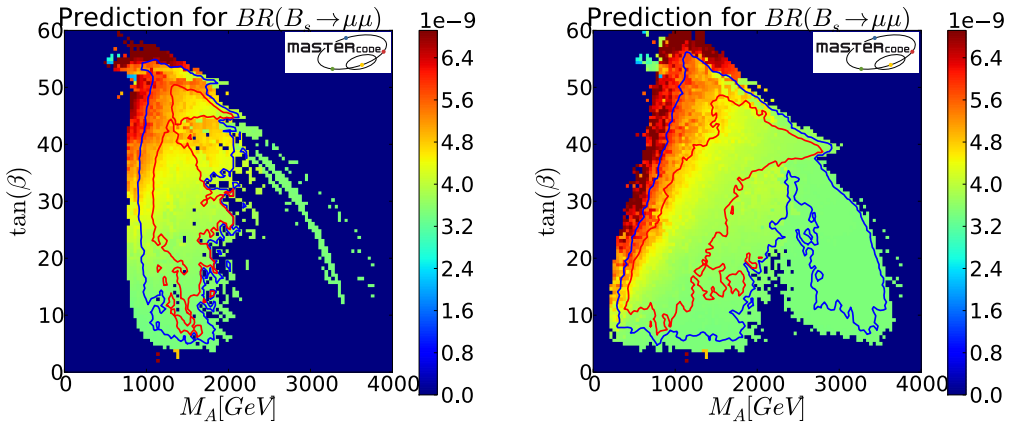


Fig. 1.10. *Prediction of the $B_s^0 \to \mu^+\mu^-$ branching ratio in the $M_A - \tan\beta$ plane for CMSSM (left) and NUHM1 (right) models. The red (blue) contour shows the 68 % (95 %) CL exclusion obtained including the results from LHCb, CMS, ATLAS and CDF available in 2012 [45].*

Higgs masses [46, 47]. The plots are not updated with the latest measurements (the constraints have been obtained using a combination of LHCb, CMS, ATLAS and CDF results available in 2012), but they give an idea of the power of the $B_s^0 \to \mu^+\mu^-$ channel in constraining all these theories beyond the SM.

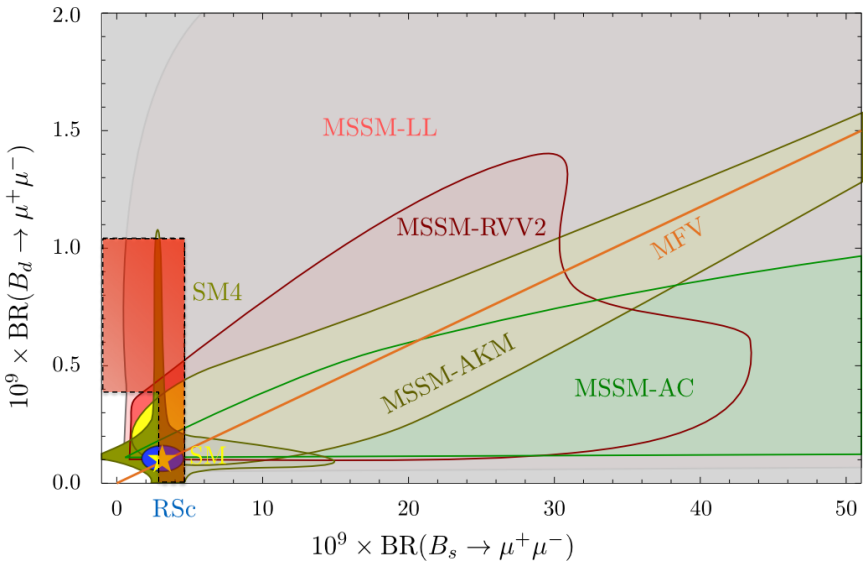


Fig. 1.11. *Correlation between the branching ratios of $B_s^0 \to \mu^+\mu^-$ and $B_d \to \mu^+\mu^-$ in MFV, SM4 and four SUSY flavour models (see text) [50]. The grey area is excluded by the experiments (2012). The red area has been added to the original plot and takes into account of the latest combination presented by CMS and LHCb [28].*

Finally, figure 1.11 [50] shows the correlation between $B_s^0 \to \mu^+\mu^-$ and $B_d \to \mu^+\mu^-$ in several physics models like MVF, the Randall-Sundrum model with custodial protection (RSc) [48], the Standard Model with a sequential fourth generation (SM4) [49] and four SUSY flavour models (MSSM-AC, MSSM-AKM, MSSM-RVV2, MSSM-LL)[12]. The grey area shows the parameter space of the supersymmetric models (in which $\tan\beta$ can be large) that has been ruled out by the experimental results. The additional red area has been added considering the latest measurements from CMS and LHCb, leading to a much more constrained situation. However, the models like SM4, RSc or SUSY with small $\tan\beta$, in which the NP in the $B_s^0 \to \mu^+\mu^-$ channel enters via the operators $\mathcal{O}_{10A}$ (in equation 1.32), are starting to be probed only now.

[12]The acronyms stand for the models by Agashe and Carone (AC)[51], Ross, Velasco-Sevilla and Vives (RVV2)[52], Antusch, King and Malinsky (AKM)[53] and a model with left-handed currents only (LL)[54]. More details can also be found in this analysis [55].

Chapter 2

The Large Hadron Collider and the ATLAS Experiment

The Large Hadron Collider (LHC) at CERN is extending the frontiers of particle physics with its unprecedented high energy and luminosity. It is designed to collide bunches of up to 10^{11} protons 40 million times per second with 14 TeV centre-of-mass (c.m.) beam energy and target luminosity of 10^{34} cm^{-2} s^{-1} [56]. It can also collide heavy ions with an energy of 2.8 TeV per nucleon and a luminosity of 10^{27} cm^{-2} s^{-1}.

After the incident in September 2008 caused by a faulty electrical connection between two of the accelerator's magnets that resulted in mechanical damage and release of helium from the magnet cold mass into the tunnel [57], LHC restarted colliding low energy beams at the end of 2009. 2010 was mostly dedicated to machine commissioning and establishing confidence in the critical machine protection system with a c.m. energy of 7 TeV and a peak luminosity of $2.1 \cdot 10^{32}$ cm^{-2} s^{-1}. Throughout all 2011 the luminosity was increased (keeping the beam energy at 3.5 TeV) reaching values up to $2.4 \cdot 10^{33}$ cm^{-2} s^{-1}. 2012 was the "production year" where the c.m. energy was raised to 8 TeV and the luminosity reached up to $7.7 \cdot 10^{33}$ cm^{-2} s^{-1}. One of the main features of operations in 2011 and 2012 was the use of high bunch intensity with 50 ns bunch spacing. This gave a good luminosity, but at the cost of high pile-up[1] in the experiments (in ATLAS and CMS the peak number of interaction per bunch crossing in the three years of data taking was around 4, 17 and 37 respectively). After the first long technical stop, LHC will restart circulating beams in early 2015 at 13 TeV c.m. energy, while collisions at 50 ns bunch spacing will be available in June. Finally, in August LHC is planned to start colliding beams at 25 ns with an expected peak luminosity of $1.6 \cdot 10^{34}$ cm^{-2}s^{-1} (the LHC 2015 schedule can be found in [58]).

ATLAS (A Toroidal LHC ApparatuS) [59] is a multipurpose experiment designed to achieve the highest possible flexibility studying high energy physics. It operates at the LHC and is devoted to the study of interactions produced in the collision of

[1] The "pile-up" is the average number of inelastic events per crossing. It is expressed as $<N_{\text{pile-up}}> = \sigma_{inel} \cdot \mathcal{L} \cdot \tau_b$, where σ_{inel} is the inelastic cross-section, $\mathcal{L}$ is the integrated luminosity and τ_b is the time bunch spacing.

two proton beams. Since it is possible that unexpected phenomena will manifest, the ATLAS main design requirement is to be capable of performing precise measurements not only on expected physical event topologies, but also on unforeseen channels. For this reason, the detector design has to avoid biases from the physical program and from theoretical expectations. ATLAS must be capable of good identification and kinematical measurements of all the stable particles that can be produced in *pp* collisions at the LHC.

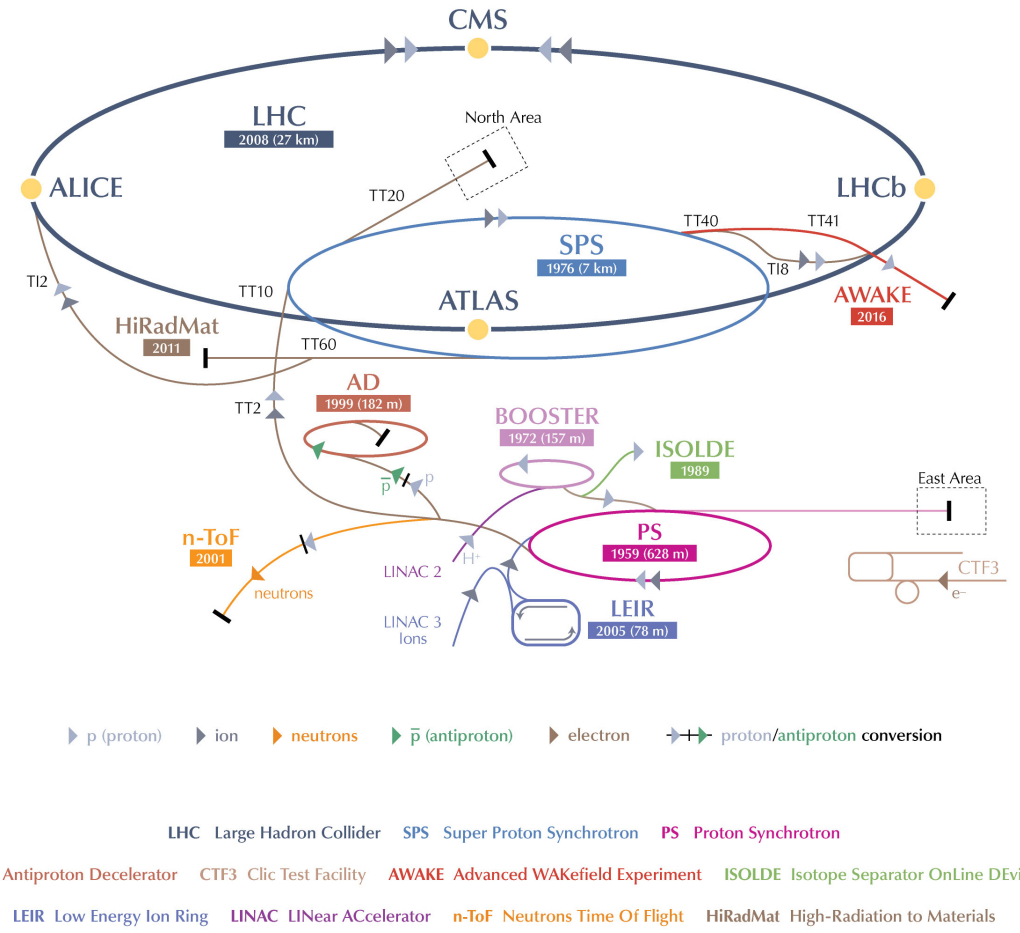


Fig. 2.1. *CERN's accelerator complex [OPEN-PHO-ACCEL-2013-056-1].*

2.1 The Large Hadron Collider

The LHC [60] is two-ring superconducting hadron accelerator and collider installed in the 26.7 km tunnel constructed between 1985 and 1989 for the LEP machine (figure 2.1 shows CERN's accelerator complex). The decision of using the LEP tunnel and the

existing injector chain (comprising linac, booster, PS and SPS) was strongly driven by cost saving. The LHC project was approved by the CERN council in December 1994. At that time, the original plan was to start with a c.m. energy of 10 TeV (to be upgraded later to 14 TeV), but after intense negotiations, in 1996 the CERN Council approved the construction of a 14 TeV machine in a single stage. The construction of colliders such as LEP was only made possible by what Nicola Cabibbo (at a Workshop held in Frascati in 1984) called the *exo-geographical transition*, i.e. the construction of machines deep under properties which does not belong to the laboratory concerned. Without this agreement, Europe could not have maintained its leading position in accelerator and particle physics.

In the next sections we will briefly review the main characteristics and performance of the LHC machine.

2.1.1 Machine Layout

The LHC tunnel has eight straight sections and eight arcs and lies between 45 m and 170 m below the surface on a plane inclined at 1.4 % sloping towards the Léman lake. Each straight section is approximately 528 m long. Four of the straight sections house the LHC detectors whilst the other four are used for machine utilities, radio frequency systems, collimation and beam abort. The two general purpose detectors are located at diametrically-opposite straight sections (see figure 2.2, left). The ATLAS detector is located at Point 1 and CMS [61] at Point 5, which also incorporates the small angle scattering experiment TOTEM [62]. Two more detectors are located at Point 2 (ALICE [63]) and at Point 8 (LHCb [64]). At these points the injection systems for the two rings are also placed. The beams only cross from one ring to the other at these four locations where bunches with nominal 25 ns separation collide simultaneously (the bunch structure is generated in the 25 GeV Proton Synchrotron (PS)). The straight sections at point 3 and 7 contain two collimation systems respectively to capture off-momentum particles (momentum collimation) and to reduce the beam halo (betatron collimation). Point 4 contains the Radio Frequency (RF) system (section 2.1.4) while Point 6 the two beam abort systems that safely extract the beams and dump them into external absorbers.

Two transfer tunnels of approximately 2.5 km connect the LHC to the CERN accelerator complex that acts as injector. The beams are injected from the Super Proton Synchrotron (SPS) at an energy of 450 GeV and, once both rings are filled, the energy is increased up to 7 TeV in about 30 minutes. The magnetic field necessary to circulate the proton beams in the ring, at their maximum energy, is around 8.3 T and can be achieved using NbTi superconducting cables cooled down to a temperature of 1.9 K (see section 2.1.3). The cryogenic equipment needed to produce about 100 tons of superfluid helium is unprecedented in scale and complexity [56].

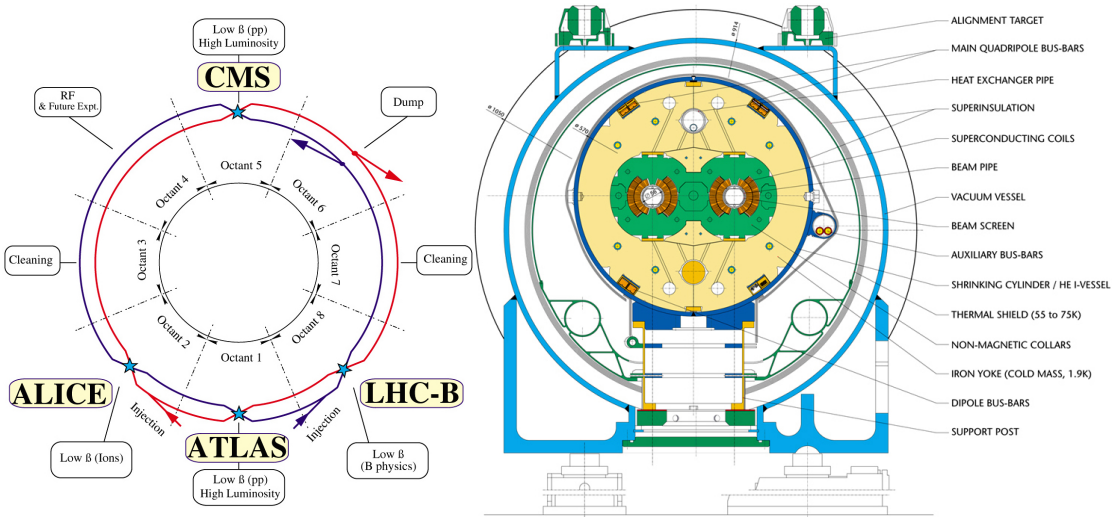


Fig. 2.2. *LHC layout (left) [LHC-PHO-1997-060-1] and cross section of a dipole (right) [LHC-PHO-1999-172-1].*

2.1.2 Machine Performance

The instantaneous power loss by a charged particle of energy E, moving uniformly on a circular orbit, is expressed as

$$P[\mathrm{W}] = \frac{e^2\,\gamma^4\,\beta^4 c}{6\pi\,\epsilon_0\,\rho^2} \xrightarrow[(\beta\to 1)]{} \frac{e^2\,c}{6\pi\,\epsilon_0}\cdot\frac{1}{\rho^2}\left(\frac{E}{mc^2}\right)^4 = 4.6\cdot 10^{-20}\cdot\frac{1}{\rho^2}\left(\frac{E}{mc^2}\right)^4 \qquad (2.1)$$

where e is the electric charge, ϵ_0 the vacuum permittivity, ρ is the bending radius (in meters), c the speed of light and m the mass of the particle. The corresponding synchrotron radiation energy emitted per turn is

$$\Delta E[\mathrm{GeV}] = P\cdot\Delta t = \frac{e^2\,\gamma^4\,\beta^4 c}{6\pi\,\epsilon_0\,\rho^2}\cdot\frac{2\pi\rho}{\beta\cdot c} \xrightarrow[(\beta\to 1)]{} = 6.0\cdot 10^{-18}\cdot\frac{1}{\rho}\left(\frac{E}{mc^2}\right)^4 . \qquad (2.2)$$

Since the energy loss per turn goes as the fourth power of E, the need to explore higher energies made impossible to continue using electrons. At the maximum c.m. energy of LEP ($\sqrt{s} \simeq 200$ GeV) the instantaneous power loss of an electron was $\sim 7.7\cdot 10^{-6}$ W (the LEP bending radius in the dipoles was 3,096 m [65]), while the energy emitted per turn via synchrotron radiation was about 3.1 GeV. Considering a beam made of 4 bunches containing roughly $1.5\cdot 10^{11}$ particles, the instantaneous emission per beam was around 4.6 MW. The eight straight sections were designed exactly to host the LEP RF cavities that compensated these high synchrotron radiation losses. The acceleration of an electron at the LHC would produce a total power loss per beam (2808 bunches, $\sim 10^{11}$ particles per bunch, 2804 m of bending radius) of $\sim 6.3\cdot 10^{10}$ MW! In order to have a reasonable emission ($\mathcal{O}(1\,\mathrm{MW})$ as for LEP) the bending radius should be about $7\cdot 10^5$ km.

Due to the small tunnel internal diameter (3.7 m), the installation of two separate proton rings would have been impossible. Therefore, the limited space available lead to the adoption of the twin-bore magnet design [66], proposed by John Blewett at the Brookhaven laboratory in 1971. The high intensity of the beam does not allow using anti-proton, and so excludes the particle-anti-particle collider configuration of a common vacuum and magnet system for both circulating beams, as used, for example, in the Tevatron [67–69]. Table 2.1 summarises the main design parameters of LHC.

The most important figure of merit for colliders like LHC is the luminosity. The number of events N_α for a certain process α produced in pp collisions is given by:

$$N_\alpha = \mathcal{L} \cdot \sigma_\alpha \tag{2.3}$$

where $\mathcal{L}$ is the luminosity and σ_α is the related cross-section. The luminosity depends on the beam parameters (summarised in table 2.1) and can be written as

$$\mathcal{L} = F \cdot \frac{f\,\gamma_r\,n_b\,N_b^2}{4\pi\,\epsilon_n\,\beta^*} \tag{2.4}$$

where f is the revolution frequency, n_b is the number of bunches, N_b is the number of protons per bunch, γ_r is the gamma relativistic factor, ϵ_n the normalised transverse beam emittance[2] and β^* is the beta function at the collision point[3]. The geometric luminosity reduction factor, F, that takes into account that the crossing angle at the interaction point is not zero, is defined as

$$F = 1 \Bigg/ \sqrt{1 + \left(\frac{\theta_c\,\sigma_z}{2\,\sigma^*}\right)^2} \tag{2.5}$$

where θ_c is the full crossing angle at the IP (285 μrad at IP1 and IP5 as nominal design), σ_z is the RMS bunch length, and σ^* the transverse RMS beam size at the IP (the above formula assumes two round beams, $\sigma_z \ll \beta$, with exactly the same parameters). Therefore, the luminosity of a collider can be increased maximising the total beam current ($f\,n_b\,N_b$), the brightness (N_b/ϵ_n) and the energy (γ_r), and minimising the beta function at the interaction point.

During nominal LHC operation the instantaneous luminosity exponentially decreases as

$$\mathcal{L} = \mathcal{L}_0 \cdot e^{-t/\tau} \quad \tau \approx 14 \text{ h} \tag{2.6}$$

[2]The particle emittance ϵ is a measure of the average spread of particle coordinate in position and momentum phase-space. The normalised emittance is defined as $\epsilon_n = \beta\gamma_r\epsilon$, where $\beta = v/c$ and γ_r is the relativistic factor. The normalised emittance does not change as function of the energy and allows to track the beam degradation.

[3]The beta function is a function related to the transverse size of the particle beam at the location s along the nominal beam trajectory. It is related to the transverse beam size as $\sigma(s) = \sqrt{\epsilon \cdot \beta(s)}$. The beam is assumed to have a Gaussian shape in the transverse direction with a width equal to $\sigma(s)$. The value of the beta function at the interaction point is named β^* that is also referred as the distance from the focus point where the beam width is twice as wide as at focus point.

		Injection	Collisions
Geometry			
Circumference	[m]	26658	
Ring separation in arcs	[mm]	194	
Beam Data			
Proton energy	[GeV]	450	7000
Relativistic gamma		479.6	7461
Number of particles per bunch		$1.15 \cdot 10^{11}$	
Number of bunches		2808	
Transverse normalised emittance	[μm rad]	3.5	3.75
Time bunch spacing	[ns]	24.95	
Spatial bunch spacing	[cm]	7.48	
Beam crossing angle	[μrad]	-	285
Circulating beam current	[A]	0.582	
Stored energy per beam	[MJ]	23.3	362
Proton energy loss per turn	[eV]	$1.15 \cdot 10^{-1}$	$6.71 \cdot 10^{3}$
Main Magnet			
Number of main bends		1232	
Length of main bends	[m]	14.3	
Field of main bends	[m]	0.535	8.33
Bending radius	[m]	2803.95	
RF System			
Revolution frequency	[kHz]	11.245	
RF frequency	[MHz]	400.8	
Harmonic number		35640	
Total RF voltage	[MV]	8	16
Synchrotron frequency	[Hz]	61.8	21.4
Peak Luminosity Related Data			
Peak luminosity in IP1 and IP5	[$cm^{-2}s^{-1}$]	-	$1 \cdot 10^{34}$
RMS bunch length (σ_z)	[cm]	11.24	7.55
RMS beam size in IP1 and IP5 (σ^*)	[μm]	375.2	16.7
β at IP1 and IP5 (β^*)	[m]	18	0.55
Luminosity reduction factor F		-	0.836
Luminosity lifetime (p-p, rest-gas, IBS)	[h]	-	14.9

Table 2.1. *LHC design parameters [60].*

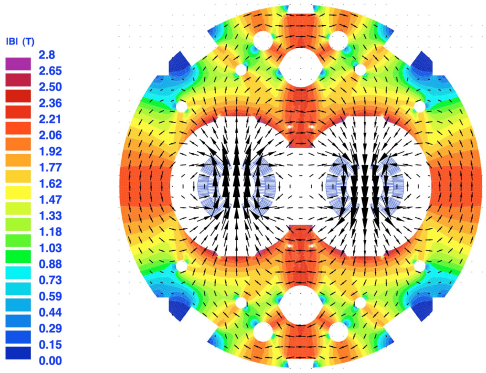


Fig. 2.3. *Magnetic flux in the LHC dipole [70].*

mainly due to beam loss from collisions. Other contributions to beam losses come from Touschek effect [4] (that results in both particles be ejected from the beam) and from particle losses due to a slow emittance blow-up. The latter can be caused by the scattering of particles on residual gas, the nonlinear force of the beam-beam interaction, RF noise and Intra-Beam Scattering (IBS) effects [5] that cause the beam size to grow (conversely to the Touschek effect, the IBS increases during the run).

2.1.3 Magnets

LHC contains more than 7000 superconducting magnets ranging from the 14.3 m long dipoles to the 10 cm long octupole/decapole correctors and 154 normal conducting "warm" magnets[6] with laminated steel cores and copper coils (about 12 % of the total number). In addition, more than 500 warm magnets (6.3 m core length, 25 mm gap height and nominal field of 1.81 T at a current of 5270 A) are installed along the two 2.8-km transfer lines (TI 2 and TI 8) between the SPS and the LHC. The LHC magnet system mainly uses NbTi Rutherford cables, cooled to a temperature of 1.9 K by means of superfluid helium, that allow to reach a field above 8 T.

The LHC ring accommodates 1232 main dipoles: 1104 in the arc and 128 in the

[4] The Touschek effect is a loss mechanism driven by large-angle Coulomb collisions in the bunch that lead to momentum transfers (enhanced by the relativistic factor) in the longitudinal plane. If the momentum deviation exceeds the RF bucket (area in RF phase where particles oscillate about the synchronous phase angle), the particle is lost. This effect is lifetime dependent and thus it decreases during the run.

[5] The intra-beam scattering is a multiple Coulomb scattering of charge particles in a beam. It causes small changes in the momentum of the colliding particles leading to an increase of the size of the beam that causes a faster decay of the luminosity.

[6] In the eight long straight sections and on each side of the four experiments, the magnetic intensity required to bend the proton beams is not as high as elsewhere and the magnetic field delivered by warm magnets is sufficient. Another advantage of warm magnets is their robustness when exposed to radiation like the one produced in the four experiments or in the long straight sections LLS3 and LLS7, where collimators clean the beam by removing particles that are located far from the central distribution of the proton bunches.

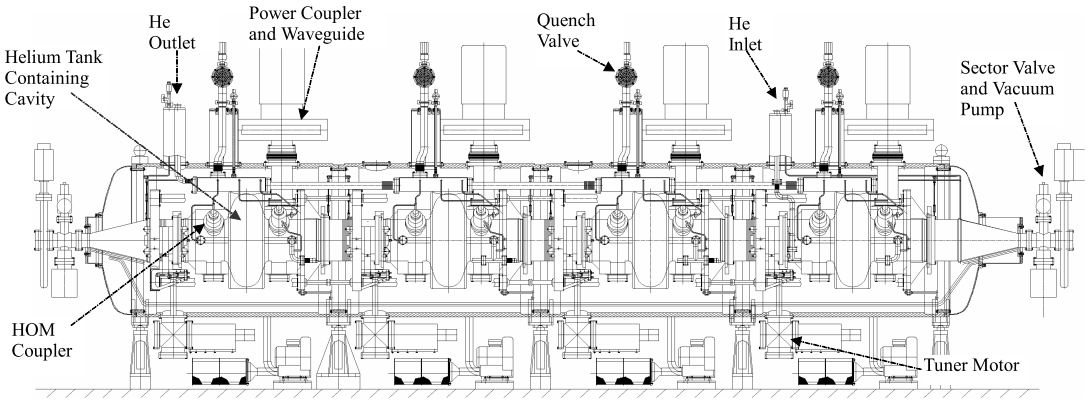


Fig. 2.4. *Four-cavity cryomodule [56].*

Dispersion Suppression (DS) regions[7]. The dipole magnets consist of two dipoles in a common iron yoke mades of low carbon steel which carries the magnetic flux, while the two coils are contained in an austenitic steel collar with low permeability. Due to the very high energy stored in the magnet system ($\sim$ 500 kJ/m), an efficient quenching protection is required. In case of a quenching, the whole coil is made resistive by firing a capacitor bank into resistive strips built into the coil and then the current is diverted through a diode until the power is completely switched off. The cross-section of the LHC dipole is reported in figure 2.2 (right), while figure 2.3 shows the magnetic flux in the dipole.

2.1.4 Radio Frequency Acceleration System

The RF system is located at Point 4 and is made of two independent set of cavities operating at 400 MHz (twice the frequency of the SPS) that provide at least 16 MV during collisions, whilst at injection 8 MV is needed. Each cavity is driven by an independent RF system, with independent klystron, circulator and load. The final design uses 8 single cell cavities per beam, each providing 2 MV (corresponding to a gradient of 5.5 MV/m), that minimises the power carried by the RF window. The cavities are grouped into two modules per beam, each containing four cells. Four cavities, each equipped with their helium tank, tuner, power coupler, HOM dampers and couplers, are grouped together in a single cryomodule (see figure 2.4). Since all cavities are identical and they can be installed in any position, if a problem raises, the cavity can be "easily" replaced.

The frequency is close to that of LEP, 352 MHz, and that allows to use the same proven technology. The cavities are made from copper on which a thin film of a few microns of Niobium is sputtered onto the internal surface. This solution has the

[7]The connection between the arcs and the straight sections is made in a transition region, the dispersion suppression region, reducing the machine dispersion at the insertion. The dispersion suppressors allow to cancel the beam dispersion arising in the arcs or caused by the horizontal crossing angle at the IP.

important advantage of making the cavity insensitive to the Earth's magnetic field, avoiding special magnetic shielding (as needed for solid niobium cavities).

2.1.5 The Vacuum System

LHC has three separate vacuum systems [56]: the insulation vacuum for cryomagnets bounded externally by the room temperature vacuum vessel and internally by the Helium-II cryostat, the insulation vacuum for helium distribution, and the beam vacuum. The insulation at cryogenic temperatures, in the absence of any significant leak, is around 10^{-6} mbar. The requirements for the beam vacuum are much more stringent, driven by the required beam lifetime and background at the experiments and are expressed in terms of gas densities normalised to hydrogen, taking into account the ionisation cross section for each gas species. The equivalent hydrogen gas densities should remain below 10^{15} $H_2\,m^{-3}$ to ensure a beam lifetime from residual gas nuclear scattering in excess of 100 hours and 10^{13} $H_2\,m^{-3}$ to minimised the background at IP.

The LHC vacuum system has to ensure an adequate beam lifetime in a cryogenic system where heat input to the 1.9 K helium circuit must be minimised and where significant quantities of gas can be condensed on the vacuum chamber. The main sources of heat come from synchrotron light radiated by the beam, energy loss by nuclear scattering, energy dissipated by the development of the electron cloud and image currents of the beam.

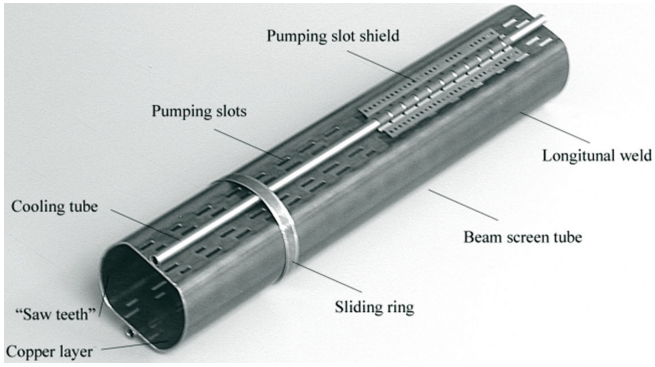


Fig. 2.5. *Picture of the beam screen, showing the flattened round shape with two cooling tubes, pumping slots and pumping slot shields [71].*

In order to remove the heat coming from all these processes, the 1.9 K cold bore of the magnets is shielded with a beam screen (nominal vertical and horizontal apertures of 34.28 mm and 44.04 respectively) cooled to between 5 and 20 K (see figure 2.5). A racetrack shape has been chosen in order to minimise the available aperture and leave space for the cooling system. Around 4 % of the beam screen surface is perforated in the flat part (the pattern has been chosen to minimise the longitudinal and transverse impedance) to allow the cold bore of the magnets at 1.9 K to act as a distributed cryopump, allowing gas to be condensed on the cold bore surface protected against desorption by bombardments with synchrotron radiation photons. A thin copper layer

(75 μm) on the inner surface of the beam screen provides a low resistance path for the image current of the beam. A saw-tooth pattern on the inner surface of this screen in the plane of bending reduces reflectivity and helps the absorption of synchrotron radiation.

2.2 The ATLAS Detector

The ATLAS detector [59] is divided into three main components (see figure 2.6). Moving outward from the interaction point, the detector consists of a tracking system to measure the directions and momenta of the charged particles, a calorimeter system that measures the energies of the electrons, photons and hadrons, and a muon spectrometer that provides the momentum and position of the muons that have enough energy to reach it. Furthermore, a system of magnets generates a magnetic field that allows the tracker and the muon spectrometer to perform measurements of the momentum of charged particles.

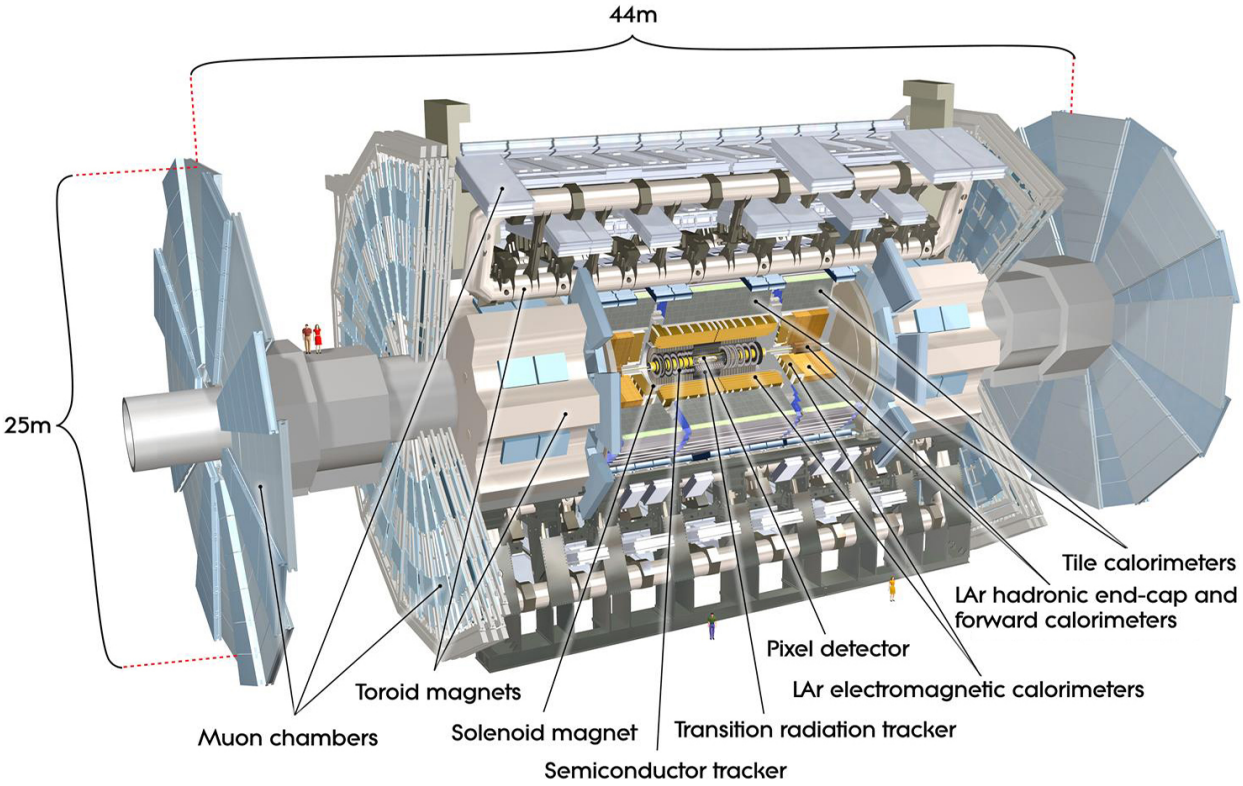


Fig. 2.6. *ATLAS detector overview [CERN-GE-0803012-01].*

2.2.1 The Coordinate System

The beam direction defines the z-axis and the x-y plane transverse to the beam, while the point of interaction is defined as the origin of the coordinate system (reported

in figure 2.7). The positive x-axis points to the centre of the LHC ring and the y-axis points upwards. Cylindrical coordinates (r, ϕ) are used in the transverse plane where ϕ is the azimuthal angle around the beam pipe. Instead of the polar angle θ, measured from the beam axis, the longitudinal direction of a particle is expressed in terms of pseudo-rapidity defined as

$$\eta = -\ln \tan\left(\frac{\theta}{2}\right) \quad , \tag{2.7}$$

corresponding to the high relativistic limit ($pc \gg mc^2$) of the rapidity $y = 1/2 \ \ln[(E + p_z)/(E - p_z)]$. The advantage of this particular definition is that differences in pseudo-rapidity are invariant under boosts along the z-axis. The transverse momentum p_T, the transverse energy E_T are defined in the x - y plane, while the Lorentz-invariant measure of the angular separation between two point objects, as observed from the origin of the ATLAS detector, is defined as

$$\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2} \quad . \tag{2.8}$$

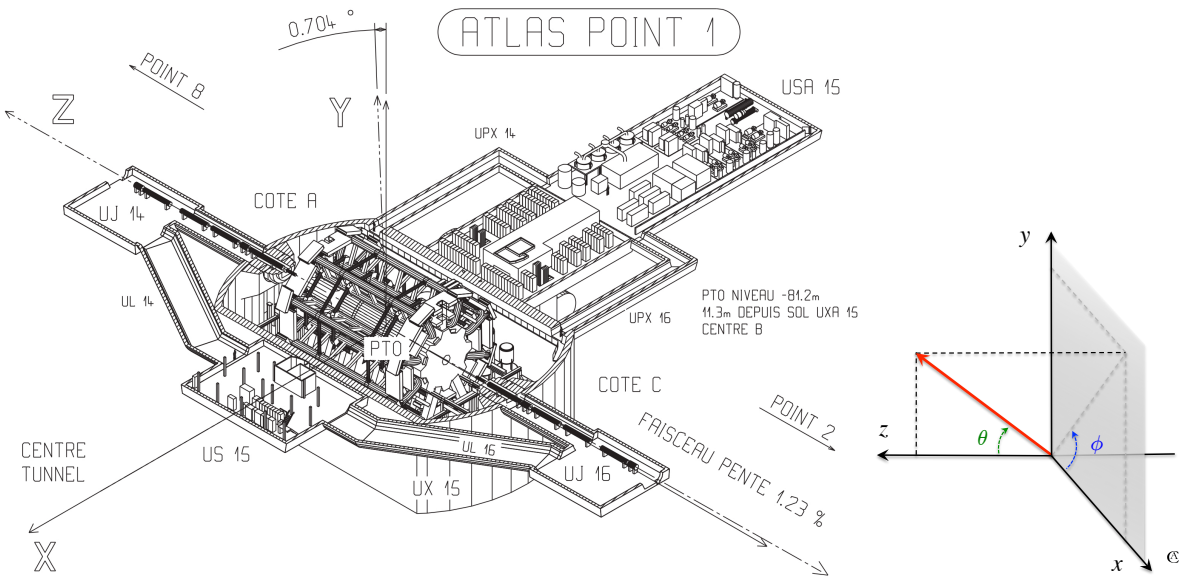


Fig. 2.7. xyz *right handed coordinate system of the ATLAS detector with* z *in the beam direction (left). Cartesian and spherical coordinate system, showing the polar angle* θ *and the azimuthal angle* ϕ *(right).*

2.2.2 The Magnet System

The ATLAS magnet system consists of one solenoid and three toroids (one in the barrel and two in the end-caps regions). It is 22 m in diameter and 26 m in length (with a stored energy of 1.6 GJ), and it is able to provide the magnetic field over a volume of approximately 12,000 m^3 (defined as the region in which the field exceeds 50 mT). The central solenoid is aligned to the beam axis and provides a 2 T axial

magnetic field for the inner detector, whereas the toroids produce a toroidal magnetic field of approximately 0.5 T and 1 T for the muons detectors in the barrel and end-cap region respectively.

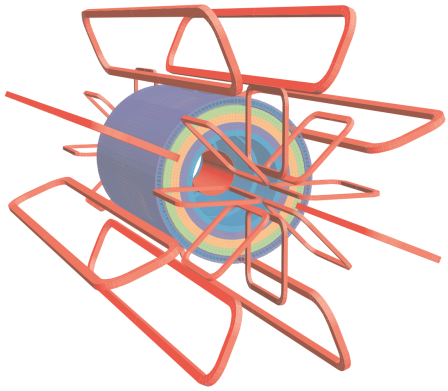


Fig. 2.8. *Geometry of the ATLAS magnet winding and the calorimeter steel [59]. The picture shows the eight barrel and end-cap toroid coils (the solenoid ending lies inside the colorimeter volume).*

The ATLAS central superconducting solenoid [72], shown in figure 2.9, has length of 5.3 m with a bore of 2.4 m and is designed to provide a 2 T axial magnetic field at the nominal operational current of 7.73 kA (the peak field in the windings is 2.6 T). Since the electromagnetic calorimeter is situated outside the solenoid, the winding should be as transparent as possible for transversing particles. To achieve a high field with the minimum thickness, the magnet is made of indirectly-cooled aluminium-stabilised superconductor. The superconducting cables consist of 12 compacted strands of multi-filamentary Nb-Ti in a copper matrix and are co-extruded in the centre of a pure aluminium stabiliser. The thickness of the cryostat is minimised by installing the solenoid between the inner wall and the cold mass of the liquid argon calorimeter that surrounds the magnet (see figure 2.9 right).

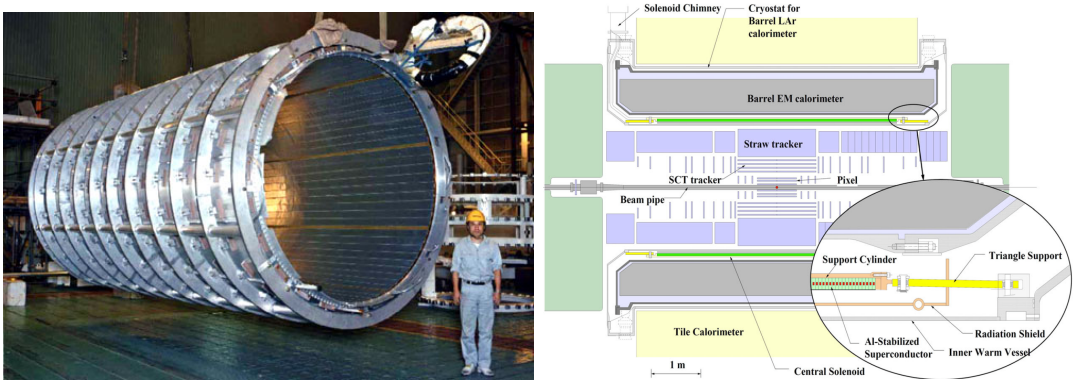


Fig. 2.9. *ATLAS solenoid after the completion of the coil winding [59] (left) and cross-section of the barrel cryostat showing the layout of the superconducting solenoid and the liquid argon calorimeter [73] (right).*

The barrel toroid consists of eight coils encased in individual racetrack-shaped,

stainless-steel vacuum vessels, that produce a tangential magnetic field of 0.5 T (peak in the winding at 3.9 T) at a nominal current of 20.5 kA (1 GJ stored energy). This provides 2 to 6 Tm of bending power in the region $|\eta| < 1.3$. The length of the barrel system is 25.3 m with an inner and outer diameter of 9.4 m and 20.1 m, respectively. The conductor and coil-winding technology is essentially the same in the barrel and end-cap toroids and it is based on winding a pure Al-stabilised Nb/Ti/Cu conductor into pancake-shaped coils, followed by vacuum impregnation.

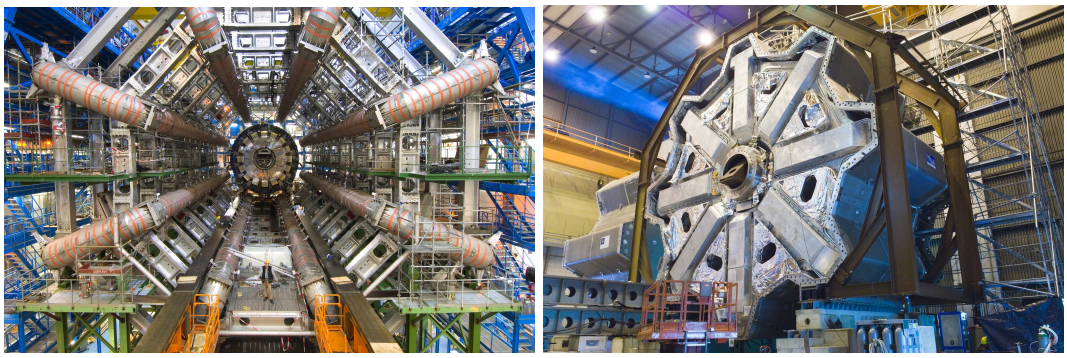


Fig. 2.10. *Barrel toroid installed in the ATLAS cavern (left) [CERN-EX-0511013-01] and end-cap toroid cold mass inserted into the cryostat (right) [CERN-EX-0611016-02].*

The two end-cap toroids (see figure 2.10 right) consist each of 8 coil modules (5×5 m^2) interlinked by 8 keystone boxes. They generate the magnetic field required for optimising the bending power in the end-cap regions of the muon spectrometer system, providing 4 to 8 Tm of bending power in the region $1.6 < |\eta| < 2.7$. The cold mass is contained in a vacuum vessel of 10.7 m diameter and 5 m length. The shape of the vessels makes it possible to insert the end-cap toroid with a 22.5 degrees turn into the barrel toroid.

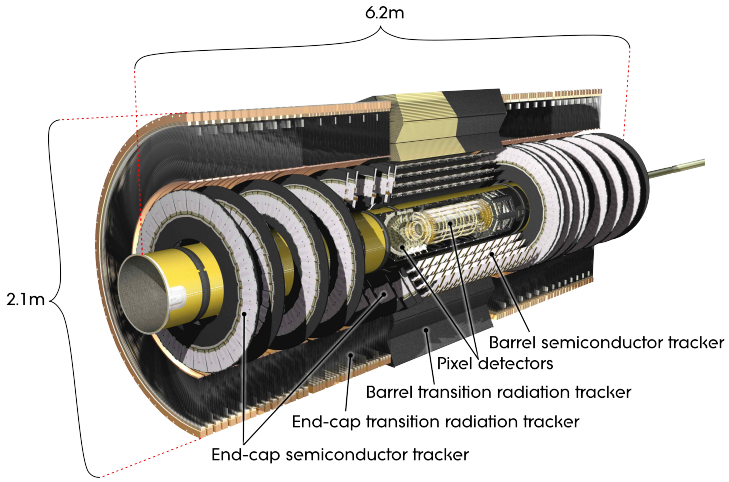


Fig. 2.11. *View of the ATLAS Inner Detector. [CERN-GE-0803014-01].*

2.2.3 The Inner Detector

The Inner Detector (ID) [59, 74] has the role to efficiently reconstruct the tracks and vertices contained in each event over the pseudo-rapidity range $|\eta| < 2.5$, and to gather information about the decay vertices of short-lived particles.

The accuracy requirements on vertices and particle momentum, necessary to achieve high precision measurements, can only be met employing detectors with high spatial precision and high granularity, given the large track density produced in the LHC pp collisions. To this purpose, the ID, completely enclosed within the central solenoid magnet, is made of high resolution detectors, placed near the beam line, and of continuous trackers, placed in the outer volume.

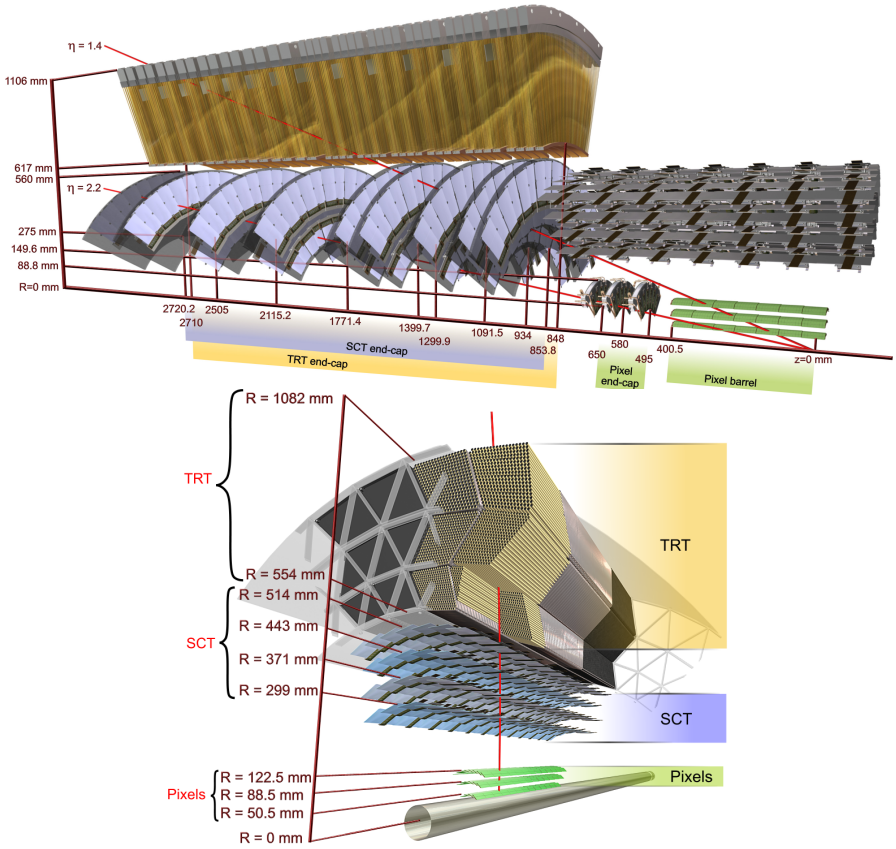


Fig. 2.12. *View of the ATLAS inner detector. Top: two charged particles with p_T of 10 GeV and η of 1.4 and 2.2 are simulated passing through the sub detectors. The tracks with $\eta = 1.4$ passes the Beryllium beam pipe, the three pixel layers, the four SCT end-caps discs and around forty end-caps TRT straw tubes. On the contrary, the particle with $\eta = 2.2$ passes through the beam pipe, only one layer and two end-caps of the pixels and the last four end-caps discs of the SCT. The TRT coverage is up to $|\eta| = 2$ [CERN-GE-0803015-05]. Bottom: one charged particle with $\eta \sim 0$ is simulated passing through the sub detectors [CERN-GE-0803014-03]*

The ID consists of three independent but complementary sub-detectors (see figures 2.11 and 2.12): the innermost, the *Pixel detector*, is composed of 50×400 μm^2 sensor

elements structured as a set of three cylinders (barrel) having their axes on the beam line and a mean radius of 50.5 mm, 88.5 mm and 122.5 mm respectively, and two sets of three discs (end-caps) placed at both ends of the barrel with a medium radius of 170 mm. In the intermediate region, the *SemiConductor Tracker* (SCT), composed of four double layers of silicon micro-strip detectors, provides, at least, four precise space points for each track. Finally, the *Transition Radiation Tracker* (TRT) uses straw tube detectors to typically guarantee about 36 spatial points for each charged track passing throughout its volume. It also provides electron identification, complementary to that of the calorimeter, over a wide range of energies. The ID is divided (see figure 2.15) into the barrel part ($|\eta| < 1.04$), the extended barrel ($1.04 < |\eta| < 1.4$) and the two end-caps ($1.4 < |\eta| < 2.5$).

2.2.3.1 Pixel Detector

The pixel detector [76] is the innermost detector of ATLAS. It consists of roughly 80 million channels and provides pattern recognition capability that is able to meet the track reconstruction requirements at the full ATLAS luminosity.

The full detector contains 1744 pixel modules (see figure 2.13), which are mounted on carbon-fibre local supports. An evaporative C_3F_8 cooling system is incorporated into the local supports to absorb the heat produced by the modules and to allow for operation at temperatures below 0° C, in order to limit the effects of radiation damage. The individual pixel modules are made of a 250 μm thick n-on-n$^+$ silicon sensor, 16 front-end chips connected throughout bump-bonding to the sensor, and a module controller chip (MCC). The sensor is divided into 47232 pixels with a size of 50 μm $\times$ 400 μm ($\sim$ 10 % of the sensors have size 50 μm $\times$ 600 μm to bridge the gaps between the readout chips in the pixel long direction and another $\sim$ 2.5 % of the electronic channels have two sensor pixels connected to bridge the gaps in the other direction).

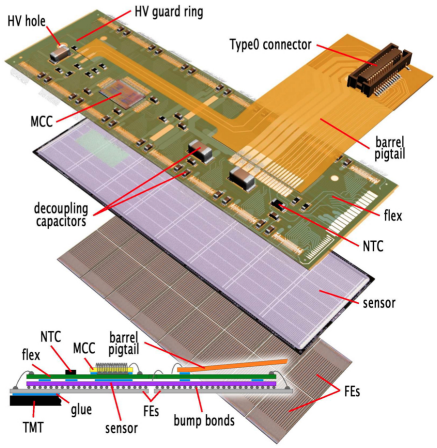


Fig. 2.13. *Elements of a pixel barrel module [59].*

Every pixel is read out through an amplifier followed by a discriminator which

detects when the pulse exceeds an adjustable threshold. The time resolution is below 25 ns as requested to associate the pixel hits with a given LHC beam-beam collision.

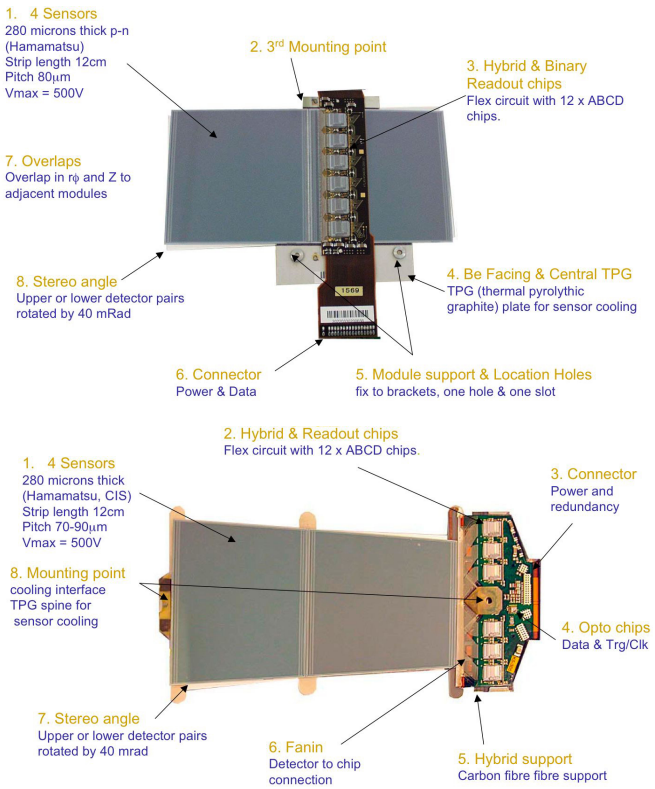


Fig. 2.14. *Barrel (top) and end-cap (bottom) SCT modules. [78]*

2.2.3.2 Semiconductor Tracker

The SCT [77] consists of four cylindrical layers (barrels) and nine disks at each end (end-caps). It comprises 4088 modules, each assembled from two pairs of single-sided silicon micro-strip sensors (64×64 mm^2, p-strips on n-type silicon), mounted back to back with a 40 mrad stereo angle. They are glued on a 380 μm-thick Thermal Pyrolytic Graphite (TPG) base-board which provides thermal and mechanical structure. A polyamide hybrid with a carbon-fibre substrate bridges the sensors on each side. Each sensor contains 768 strips at a pitch of 80 μm. The barrel contains 2020 modules made of rectangular silicon strips sensors with crystal orientation along the crystal axis <111> plus 92 modules made of sensors with <100> crystal orientation [8] produced

[8] The various crystal planes are defined using Miller indices i, l, m in the basis of the reciprocal lattice vectors, determined finding the intercepts of the plane with the three basis axes in terms of lattice constants and then taking the reciprocal of these numbers (the result is express as $(i\,l\,m)$). The notation $<i\,l\,m>$, used in this thesis, defines a set of equivalent directions in the basis of the direct lattice.

by Hamamatsu Photonics[9]. The end-caps consist of 1976 wedge-shaped and tapered strips produced by both Hamamatsu Photonics and CiS. Figure 2.14 shows the barrel and end-cap modules, while figure 2.15 shows one quadrant of the SCT.

In the SCT, most of the signal is generated by the holes drifting towards the p-implant strips. The signal from each strip is amplified and compared with a threshold in the radiation-hard front-end ABCD chip [79] (a common discriminator threshold corresponding to a charge of 1 fC is used during data taking). For each strip, the discriminator output is sampled for 3 consecutive time bins: for the triggered bunch crossing (BC) as well as the BC before and after the triggered BC. The chip is configured to register a "hit" depending on the pattern of the three sampled time bins. For cosmic ray running and for optimising the timing, a hit simply requires a signal in any of the 3 bins. For moderate luminosity (used up to 2013), the chip is configured for X1X, meaning a signal in at least the triggered time bin. At very high luminosity, the chips will require 01X, whereby signals in the BC before the trigger are vetoed. The relative occupancies of the three time bins are used to optimise the timing of the SCT, so that all hits match the 01X pattern.

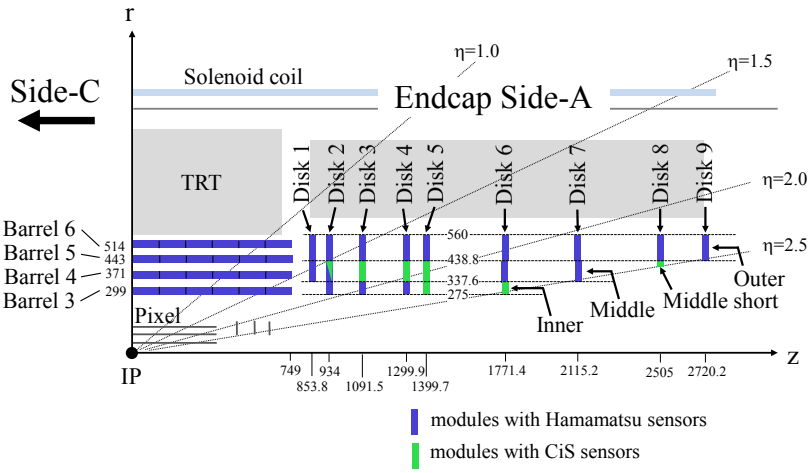


Fig. 2.15. *Schematic view of one quadrant of the SCT [77]. The numbering scheme for barrel layers and end-cap disks is indicated, together with the radial and longitudinal coordinates in millimetres.*

More details on the SCT detector can be found in chapters 3 and 4.

2.2.3.3 Transition Radiation Tracker

The ATLAS Transition Radiation Tracker (TRT) [80] is the outermost of the three tracking subsystems of the ID (see figures 2.11 and 2.12). The TRT is a straw-tube tracker that contains around 300,000 thin-walled proportional-mode drift tubes providing on average 35 two-dimensional space points with a resolution of approximately

[9]Both type of sensors were tested during the development phase, but the <111> were chosen due to easier availability. In any case, the already produced <100> sensors were included in the SCT barrel.

130 μm for charged particle tracks with $|\eta| < 2$ and $p_T > 0.5$ GeV. The larger (with respect to the silicon trackers) single-point resolution is compensated by the large number of hits per tracks and the long lever arm that makes the TRT complementary to the silicon-based tracking devices.

The drift tubes have a diameter of 4 mm and are made from wound Kapton reinforced with thin carbon fibre bundles. Each tube is filled with a gas mixture of 70 % Xe, 27 % CO_2 and 3 % O_2, and has a gold-plated tungsten wire of 31 μm diameter that is kept at ground potential. Since the wall is kept at a voltage of -1.5 kV, each tubes acts as a small proportional counter (the TRT can also provide timing information at the nanosecond level, using the high sampling frequency of the wire signals). When a charged particle crosses the TRT, it ionises the gas inside the straws, generating free electrons that drift towards the wire where they are amplified and read out. The barrel region contains 52,544 straw tubes of 1.5 m length, parallel to the beam axis, that cover a region with $|\eta| < 1$ and a radius from 0.5 m to 1.1 m, and are read out at both ends of the straws. Each side of the end-cap region contains 122,880 radial 0.4 m long straws perpendicular to the beam axis, covering the range 0.8 m $< |z| <$ 2.7 m and $1 < |\eta| < 2$ (the end-cap straws are read out at their outer end).

One of the peculiarity of the TRT is to provide electron identification using the transition radiation emitted by highly relativistic charged particles. The spaces between the straws are filled with polymer fibres (barrel) and foils (end-caps) that can produce transition radiation when a particle crosses the material boundary. The effect depends on the relativistic factor $\gamma = E/m$ and it is strongest for electrons. The soft (5-30 keV) X-rays that are produced can be absorbed by the Xe atoms, depositing additional energy in the gas and generating higher read-out signals.

2.2.4 Calorimetry

The calorimeters [81, 82] play a key role in the reconstruction of various interesting physical channels. Since they increase their intrinsic resolution as the measured energy grows (unlike other detectors such as the magnetic spectrometers), they are therefore particularly suitable for very high energy physics applications.

The main tasks of the ATLAS calorimetry system are: precise measurement of energy, position and shower shape for electrons, photons and jets; estimation of the missing transverse energy; particle identification, separating electrons and photons from hadrons and jets, and hadronic τ decays from background jets; event selection for triggering purpose. The calorimeters also limit the punch-through into the muon system.

The ATLAS calorimetry system surrounds the ID and is, in its turn, enclosed inside the muon spectrometer. This system comprises ElectroMagnetic (EM) calorimeters, for precision measurements of electrons and photons, and a larger tile hadronic calorimeter (with coarser granularity) placed in the outer and end-cap regions (see figure 2.16).

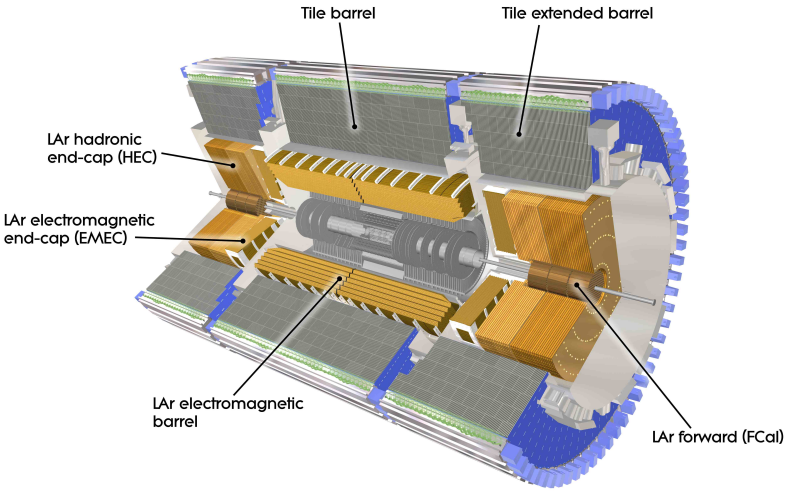


Fig. 2.16. *The ATLAS calorimetry system [CERN-GE-0803015-01].*

2.2.4.1 LAr electromagnetic Calorimeter

The EM calorimeter is a lead-Liquid Argon (LAr) detector with accordion-shaped kapton electrodes and lead absorber plates over its full coverage and it is divided into a barrel part ($|\eta| < 1.475$) and two end-caps ($1.375 < |\eta| < 3.2$), each housed in their own cryostat. The barrel calorimeter consists of two identical half-barrels, separated by a small gap of 4 mm at z = 0. Each end-cap calorimeter is mechanically divided into two coaxial wheels, an outer wheel covering the region $1.375 < |\eta| < 2.5$, and an inner wheel covering the region $2.5 < |\eta| < 3.2$.

2.2.4.2 Hadronic Calorimeter

The hadronic calorimeter is composed of the tile calorimeter, the LAr-hadronic end-cap calorimeter and the LAr forward calorimeter. The tile calorimeter is a sampling calorimeter, placed outside the EM calorimeter envelope, that uses steel as the absorber and scintillating tiles as the active material. The barrel covers a region with $|\eta| < 1$, while the two extended barrels the intermediate region with $0.8 < |\eta| < 1.7$. The LAr-Hadronic End-cap Calorimeter (HEC) consists of two independent wheels per end-cap placed behind the end-cap EM calorimeter and sharing the same LAr cryostats. The HEC covers the region: $1.5 < |\eta| < 3.2$. Finally, the LAr Forward Calorimeter (FCal), integrated into the end-cap cryostats, consists of three modules in each end-cap: the first, made of copper, is optimised for electromagnetic measurements, while the other two, made of tungsten, measure predominantly the energy of hadronic interactions (coverage: $3.1 < |\eta| < 4.9$).

2.2.5 The Muon Spectrometer

In the LHC *pp* collisions, muons are produced in many final states and used in a variety of studies such as precision measurements of SM parameters, Higgs searches,

production cross-sections, B meson decays, CP violation effects and new physics searches. For that reason, the ATLAS detector has been equipped with a muon system [83] capable of efficiently identifying momenta from a few GeV up to several TeV.

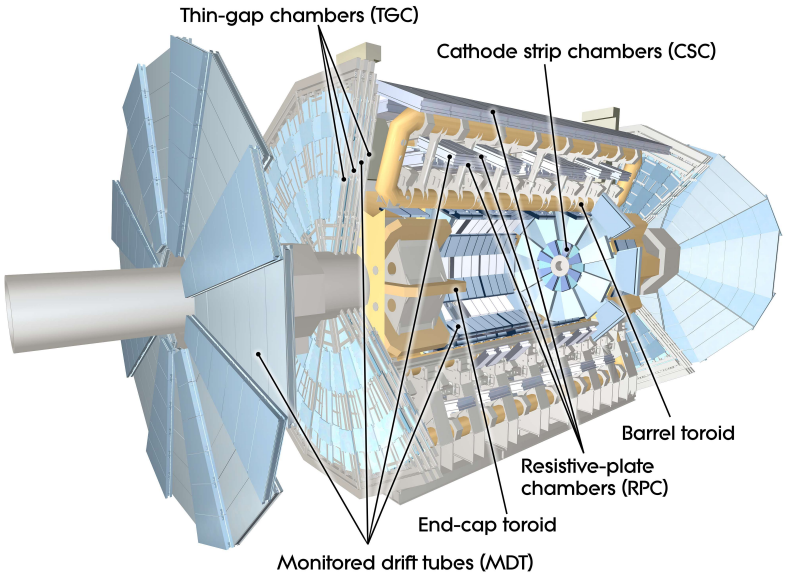


Fig. 2.17. *The ATLAS muon system [CERN-GE-0803017].*

The ATLAS Muon Spectrometer (MS) is based on the magnetic deflection of muon tracks in the three superconducting air-core toroid magnets, instrumented with separate trigger and high-precision tracking chambers (figure 2.17). As a matter of fact, only muons with a transverse momentum above 3 GeV can reach the MS by passing through the calorimeters. In the region with $|\eta| < 1.4$ the magnetic bending is provided by the large barrel toroid, for $1.6 < |\eta| < 2.7$ the particles are subjected to the field generated by the two smaller end-caps magnets inserted into both ends of the barrel toroid. Over $1.4 < |\eta| < 1.6$ (transition region) the magnetic deflection is provided by a combination of barrel and end-cap fields.

For muons with a p_T less than 100 GeV the momentum resolution can be enhanced requiring that the track reconstructed in the MS is extrapolated back to the ID and thus matched with the one reconstructed in the tracking detector. These muons are named "combined muons". Two different procedures can be used to perform this track matching; one is based on a statistical combination of the two independent track-parameter measurements; the other one performs a global fit on both the ID and the MS spatial measurements. In both cases a resolution below 2 % (for $p_T <$ 100 GeV) and around 10 % (for $p_T \simeq 1$ TeV) can be achieved. Furthermore, the matching between tracks reconstructed by different detectors enhances the rejection power against muons produced by secondary interactions as well as those coming from the "in-flight" decays of π and K mesons.

For muons with very low energies (< 3 GeV), the number of data points is not

sufficient to reconstruct a muon track from MS information only. If these data points are matching an extrapolated ID track, the candidate is considered a reconstructed muon. These muons are named “tagged muons”.

The muon spectrometer plays also a crucial role in the trigger selection. In order to be able to produce a fast decision signal, it is equipped with detector elements less precise but faster than those used for track reconstruction and momentum measurements.

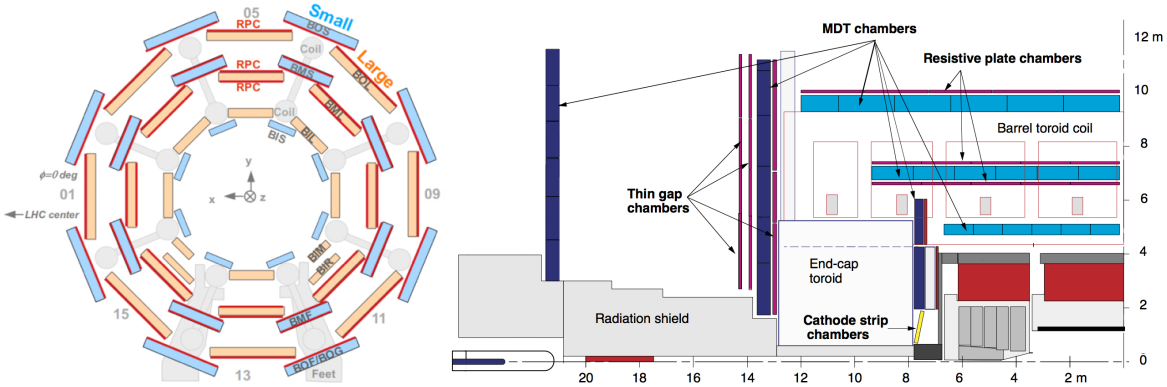


Fig. 2.18. *Cross section of the barrel muon system perpendicular to the beam axis, i.e. non-bending plane, (left) [84], and in a plane along the beam axis, i.e. bending plane, (right) [85].*

The design of the muon spectrometer is shown in figure 2.18. The barrel consists of three concentric layers at radii of about 5 (inner layer), 8 (middle) and 10 (outer) meters. Each layer consists of Monitored Drift Tube (MDT) chambers that provide precision measurements to determine the momentum. The middle and the outer layers are also equipped with Resistive Plate Chamber (RPC) that provide the barrel trigger system. A very similar layout is present in the end-caps. Three wheels of MDTs are mounted perpendicular to the beam axis at a longitudinal distance of 7.5, 14 and 22.5 m from the IP, with an exception of the innermost layer, where, close to the beam pipe, Cathode Strip Chambers (CSC) replace the MDT chambers. For the end-cap, a different trigger chamber technology, the Thin Gap Chambers (TGC), has been chosen.

2.2.5.1 Momentum Measurement Detector

The monitored drift tube chambers are made of pressurised drift tubes with a diameter of 29.97 mm, operating with Ar/CO_2 (93/7) at a pressure of 3 bar (see figure 2.19 left). When a particle crosses the tube, the electrons, produced by the ionisation of the gas, are collected at the central tungsten-rhenium wire, that has a diameter of 50 μm and is kept at a potential of 3080 V. The central conductor, holding the wire, also serves for the gas transfer in and out of the tube. Signal transmission to the electronics and connection to the HV supply system are at opposite ends.

Figure 2.19 (right) shows the mechanical structure of the MDTs. The chambers are rectangular in the barrel and trapezoidal in the end-caps. The design has been studied in order to optimise the solid angle coverage, while respecting the magnet coils,

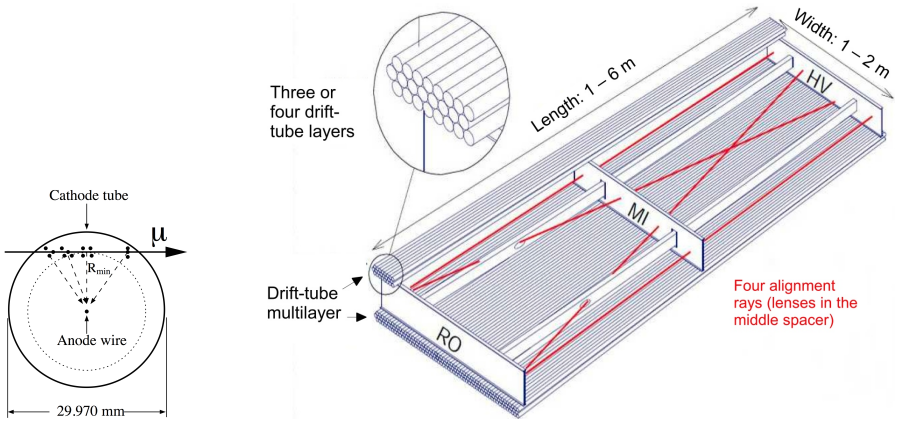


Fig. 2.19. *Cross section of a drift-tube (left) and schematic view of a MDT chamber (right) [59].*

support structures and access ducts. In both barrel and end-caps the direction of the tubes is along ϕ, i.e. the centre points of the tubes are tangential to circles around the beam axis.

The MDTs can safely operate up to a counting rate of about 150 Hz/cm^2, that is exceeded in the region $|\eta| > 2$ in the first layer of the end-cap. For that reason, the MDTs are replaced by CSCs that are able to reach rates of about 1000 Hz/cm^2. The CSCs are multi-wire proportional chambers with the wires oriented in the radial direction. Both cathodes are segmented, one with the strips perpendicular to the wires (providing the precision coordinate) and the other parallel to the wires (providing the transverse coordinate). The position of the track is obtained by interpolation between the charges induced on neighbouring cathode strips.

2.2.5.2 Trigger chambers

The trigger chambers of the muon system should be able to provide fast information on muon tracks, allowing the L1 trigger to recognise their multiplicity and approximate energy range, as well as to provide an bunch crossing identification. The trigger detectors should have an acceptance range of $|\eta| \leq 2.4$ and full ϕ coverage.

In the barrel region, Resistive Plate Chambers (RPC) are used due to good spatial and time resolution as well as adequate rate capability. They are arranged in three concentric cylindrical layers around the beam axis that are referred to as trigger stations. The RPC is a gaseous parallel electrode-plate detector made of two resistive plates (phenolic-melaminic plastic laminate) kept parallel to each other at a distance of 2 mm by insulating spacers and filled with a mixture $C_2H_2F_4$/Iso-C_4H_{10}/SF_6 (with proportions 94.7/5/0.3). The electric field between the plates of about 4.9 kV/mm allows avalanches to form along the ionising tracks towards the anode. The signal is read out via capacitive coupling to metallic strips, which are mounted on the outer faces of the resistive plates. A RPC has no wires, which simplifies its construction and makes chambers less sensitive to small deviations from planarity if appropriate spacers

are used to keep the gap width constant.

In the end-caps, Thin Gap Chambers (TGC) have been used. They work on the same principle of the multi-wire proportional chambers and are able to provide a good time resolution and a high rate capability (the resolution is mainly determined by the readout channel granularity, which can be adjusted to the need by wire ganging).

2.2.6 The ATLAS Detector Resolution

Table 2.2 summarises the energy and momentum resolution of the ATLAS detector components, while table 2.3 shows the spacial and time resolution of all ATLAS sub-detectors.

Detector component	**Required resolution**	**η coverage**	
		Measurement	**Trigger**
Tracking	$\sigma_{p_T}/p_T = 0.05\,\% p_T \oplus 1\,\%$	± 2.5	
EM calorimetry	$\sigma_E/E = 10\,\%/\sqrt{E} \oplus 0.7\,\%$	± 3.2	± 3.2
Hadronic calorimetry			
Barrel and end-cap	$\sigma_E/E = 50\,\%/\sqrt{E} \oplus 3\,\%$	± 3.2	± 3.2
Forward	$\sigma_E/E = 100\,\%/\sqrt{E} \oplus 10\,\%$	$3.1 < \lvert\eta\rvert < 4.9$	$3.1 < \lvert\eta\rvert < 4.9$
Muon spectrometer	$\sigma_{p_T}/p_T = 10\,\%$ at $p_T = 1$ TeV	± 2.7	± 2.4

Table 2.2. *Momentum and energy resolution of the ATLAS detector components. Note that, for high-p_T muons, the muon-spectrometer performance is independent of the inner-detector system. The units for E and p_T are in GeV [59].*

Detector component	**Spacial resolution**		**Time resolution**
	R - ϕ plane	z/R	
Pixel	10 μm	115 μm	-
SCT	17 μm	580 μm	-
TRT	130 μm	-	3 ns
Muon spectrometer			
MDT	-	35 μm (z)	-
CSC	5 mm	40 μm (R)	7 ns
RPC	10 mm	10 mm (z)	1.5 ns
TGC	3-7 mm	2-6 mm (R)	4 ns

Table 2.3. *Spacial and time resolution of the ATLAS detector components [59].*

2.3 The ATLAS Data Acquisition and Trigger System

The expected bunch crossing rate for the design LHC operation is 40 MHz (for 25 ns of time gap between consecutive bunches). Since, on average, the size of an event is about 1.5 MB [59], it is obvious how it would be impossible to record all data produced in *pp* collisions. In addition, not all events are relevant for the ATLAS physics program: most analyses study interactions with high energy and momentum transfer that represent a small fraction of the total *pp* cross-section. For these reasons, we need a system capable of selecting a manageable number of interesting events out of the overwhelming number of primary interactions. The maximum "to-tape" rate available is 200 Hz (400 Hz was allowed in 2012 data taking).

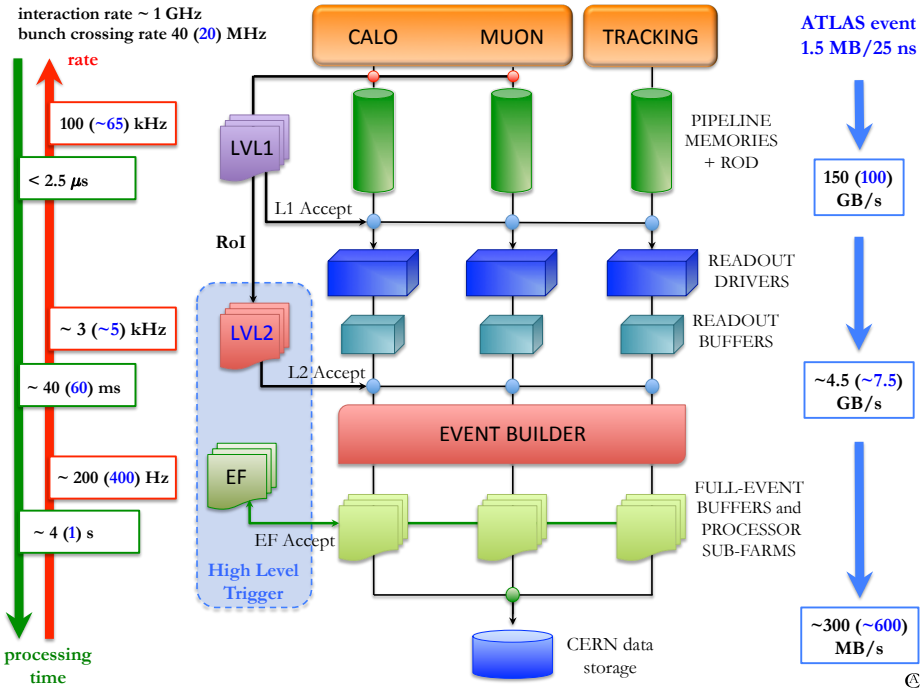


Fig. 2.20. *The ATLAS trigger and DAQ chain (the numbers in parentheses are the values of 2012 data taking period).*

The selection strategy adopted is a two-step procedure [59] [75]. The first step is the "online" (i.e. in real time) selection of interesting events, to reduce the huge bunch crossing rate to a manageable event rate which is written to mass storage devices. In the second step, the stored data are reconstructed "offline" to yield quantities like tracks, energy clusters, jets, missing transverse energy, secondary decay vertices, etc. The online system is composed of three different levels (see figure 2.20).

The *first level trigger* (L1) is designed to operate at a maximum rate of 100 kHz. The L1 decision is based on information, with a coarse granularity, from two sub-detector systems: the muon trigger chambers and the calorimeters. Quantities used in the L1 decision are typically: the estimated transverse momentum of muon

candidates (from the MS), the missing (traverse) energy, the total energy deposited and the occurrence of isolated energy depositions (from the calorimeters). Due to the very small latency time ($< 2.5\,\mu$s) the L1 trigger is implemented using custom designed electronics, and only information from the relevant sub-components is available. The detector signal outputs are temporary stored in pipe-line memories (located on the electronics front-end of each component) and they are passed to the L2 trigger only if the event is accepted by the L1 trigger. The trigger decision at L1 is made by the Central Trigger Processor (CTP) that, starting from the information of calorimeters and muon trigger processors, forms trigger conditions [10] using lookup tables. The maximum number of conditions available is 256, combined to form up to 256 trigger items, where every trigger condition may contribute to every trigger item. For events accepted by the L1, the information of all sub-detector systems is pre-processed and stored in the so-called Read-Out Buffers (ROBs). For each event, the L1 trigger also defines one or more Regions of Interest (RoIs), that are geometrical regions of the ATLAS detector where the selection process has identified interesting features. This information is used as seed for the L2 trigger.

The *second level trigger* (L2) uses both the L1 output and the data stored in the ROBs to further reduce the data rate to a maximum of 3 kHz. Even though the L2 has access to the full detector information, the selection is generally restricted to the RoIs. For a L1 muon trigger, the L2 will use the information from the precision MDT chambers to improve the muon momentum estimate, which allows a more precise selection on this quantity. For a L1 calorimeter trigger, the L2 has access to the full detector granularity, and has, in addition, the possibility to require a match with a track reconstructed in the ID. For events accepted by the L2, the data fragments stored in the ROBs are collected by the so-called Event Builder and written into the Full Event Buffers. The third trigger stage is the *Event Filter* (EF) that has access to the complete event and by means of complex offline algorithms (suitably adapted to be run in the online framework) reduces the to-tape rate below 200 Hz. The L2 and EF triggers constitute the so-called High Level Trigger (HLT).

A scaling factor (called "prescale") can also be applied to reduce the rate of a certain trigger.

Considering a maximum trigger rate of 200 Hz, the data rate written to mass storage corresponds to 300 MB/s. For a nominal LHC running year ($\sim 10^7$ s) the total data volume produced is about 3 PB ($3 \cdot 10^9$ MB).

2.3.1 The Muon Trigger

The ATLAS muon trigger system [59] selects events with muons in three steps. The first step uses fast-response trigger chambers and custom-built hardware to generate a L1 trigger based on hit coincidences. The second and third step are software based and make up the High Level Trigger (HLT).

[10] Such a condition could be, for example, that the multiplicity of a particular muon threshold has exceeded one, i.e. at least two muons in this event have passed this threshold. For this event, this trigger condition is set to "true".

2.3.1.1 The First Level Trigger

The L1 muon trigger uses information provided by dedicated fast trigger chambers of the MS, the RPCs and the TCGs (see section 2.2.5.2). It identifies candidates by a coincidence of hits in 2 layers ("low p_T") or 3 layers ("high p_T") of the trigger chambers (see figure 2.21). The p_T of the muon is estimated by looking at the degree of deviation of the hit pattern from a straight line. It is then classified according to a set of predefined thresholds indicated by labels like MU4 (for $p_T > 4$ GeV). The "low p_T" triggers are MU10 signatures and lower, while MU11 and higher are "high p_T" triggers. The efficiency of the high p_T triggers in the barrel is around 6 % lower than the low p_T due to the requirement of the third layer of coincidence.

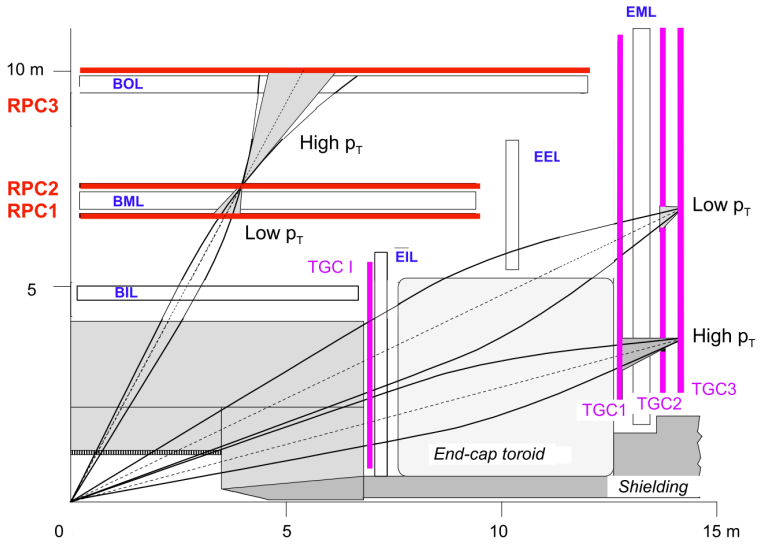


Fig. 2.21. *Quarter longitudinal section of the muon system [88]. The curved lines represent muon tracks. Low p_T triggers require a coincidence between 2 layers of trigger chambers, whereas high p_T triggers between 3 layers. BIL, BML, BOL, EIL, EEL and EML are locations of MDT chambers.*

Beside the different trigger p_T thresholds, two main configurations are used: the single muon or the di-muon trigger where, respectively, only one or two different muon signatures are required. For each trigger signature, a RoI is defined along the estimated muon direction with a size of $\Delta\eta \times \Delta\phi = 0.1$ in the barrel and $\Delta\eta \times \Delta\phi = 0.3$ in the end-caps.

2.3.1.2 The High Level Trigger

The HLT triggers, seeded by the L1 RoIs, refine the selection by accessing the data coming from the precision MS chambers, the high granularity calorimeter cells and the ID sub-components. Only information contained in the RoI (that constitutes around

2 % of the entire data) can be used by the trigger [11].

L2 and EF follow the same basic strategy: first of all they reconstruct a "standalone muon" track using the information coming from the trigger chambers and the precision chambers, then they combine it with a track reconstructed in the ID to form a "combined muon". The second strategy, that runs only at EF, starts from ID tracks around the L1 RoI and extrapolates those to the muon detectors to collect the data from the MS either to form a combined muon candidate or to just tag the ID track as a muon ("segment tagged muons"). The two strategies were running in parallel during 2011 data taking, but they were combined afterwards to reduce the processing time.

To reduce p_T thresholds of the L1 trigger, avoiding exceeding the sustainable trigger rates, in 2012 isolation was added as an extra requirement to the muon candidate. The isolation

$$\mathrm{I}^{\mu} = \left(\sum_{\mathrm{trks}} p_T^{\mathrm{trks}} \; - \; p_T^{\mu} \right) \Big/ p_T^{\mu} \tag{2.9}$$

is defined as difference between the sum of the p_T of all ID tracks (with $p_T > 1$ GeV and within a cone with $\Delta R = 0.2$ centred around the muon candidate) and the p_T of the muon, divided by the p_T of the muon itself (in order to get the relative track isolation).

2.3.2 The B-physics Triggers

The majority of the B-physics analyses need to use low p_T muons. For this reason, ATLAS developed specific di-muon triggers, with low muon p_T, that allow to keep the rate under control and, in the meantime, provide a high signal efficiency at the HLT.

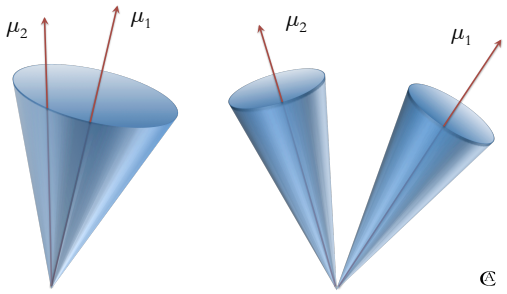


Fig. 2.22. *Schematic view of the di-muon L2 trigger algorithm seeded by one (left) or two (right) L1 trigger muons.*

At L1, the di-muon trigger requires two muon signatures that are generically labelled as `L1_2MUx` where the same p_T threshold is applied to both muons, or `L1_MUx_MUy` where two different thresholds are applied to each muon. Examples of L1 signatures

[11] During special LHC runs with low luminosity ($\sim 10^{31}$ cm^{-2}s^{-1}) the L2 triggers can access data from the whole detector.

are `L1_2MU0` [12] where no p_T cut is applied to both muons, `L1_2MU4` with $p_T^{(\mu_1,\mu_2)} > 4$ GeV and `L1_2MU4_MU6` where two muons have a p_T threshold set at 4 GeV and at 6 GeV, respectively.

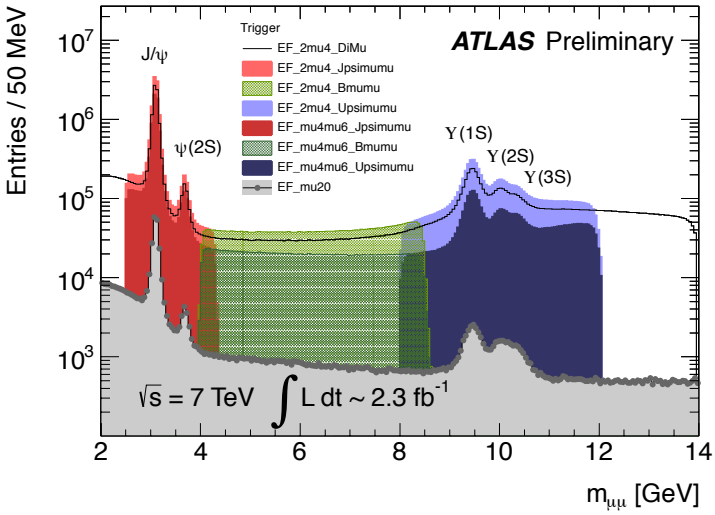


Fig. 2.23. *Invariant mass of oppositely charged muon candidate pairs selected by different triggers [89]. **EF_2mu4** is a signature with two muon triggers at L1, confirmed at the HLT with a threshold p_T cut of 4 GeV. **EF_mu4mu6** has one muon with a threshold at 4 GeV and the other at 6 GeV. **EF_mu20** has a single muon with a threshold of 20 GeV. The additional labels **_Jpsimumu**, **_Bmumu**, **_Upsimumu** and **_DiMu** denote an invariant mass window applied to the trigger objects respectively in the regions J/ψ(2.5-4.3 GeV), B_s(4-8.5 GeV), Υ(8-12 GeV) and the combined range of the three windows (1.5-14 GeV).*

Depending on the algorithm used at L2 to select the di-muon pair, the B-physics trigger can be divided in two categories. The first one, used for luminosities below 10^{33} cm^{-2} s^{-1}, is seeded by a L1 muon RoI and looks for a second muon within a wide $\Delta\eta \times \Delta\phi \simeq 0.75 \times 0.75$ region (figure 2.22, left). This is done extrapolating ID tracks to the MS. The second category (figure 2.22, right), used up to high luminosities, is constituted by the topological di-muon triggers seeded by two L1 muon RoIs, each muon is then confirmed separately at the HLT. The trigger signatures used for the L2 triggers follow the same criteria used for L1. An example is `L2_2mu4` where the lower case letters mean that a confirmation of the L1 muons (`L1_2MU4`) is performed [13].

At EF, to further reduce the trigger rate, extra requirements in opposite charge, invariant mass and vertex χ^2 match [14] are added. An example of EF trigger signature is `EF_2mu4_Jpsimumu` where two muons with a p_T above 4 GeV are requested (seeded

[12] This trigger has been used only during the first part of 2011 data taking period. With the increasing of the luminosity, a p_T threshold had to be applied to keep the rate low.

[13] This trigger can run without any pre-scale applied for luminosities $< 10^{33}$ cm^{-2} s^{-1}.

[14] A vertex finding algorithm, derived from the offline reconstruction, is used to verify the common vertex assumption.

by `L1_2MU4` and `L2_2mu4_Jpsimumu`). The extension `_Jpsimumu` in the name refers to the hypothesis of having two oppositely charged muon trigger objects, with an invariant mass in the range between 2.5 and 4.3 GeV and with a vertex χ^2 match < 20. Figure 2.23 reports the invariant mass distribution of muon candidates pairs selected by different triggers.

2.3.2.1 Muon Reconstruction and Trigger Efficiency

A tag-and-probe method has been used to determine the muon reconstruction efficiency directly from data in fine bins of muon charge-signed pseudo-rapidity and transverse momentum. The basic strategy is to select events with at least one identified and triggered combined muon, and use these tagged events to collect ID tracks (the "probes") consistent with coming from a $J/\psi \to \mu^+\mu^-$ decay. This method allows measurement of the probe muon efficiency independently from the MS system.

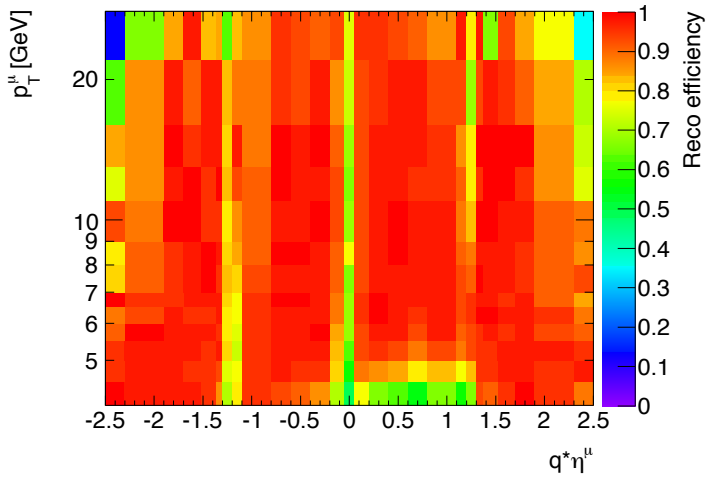


Fig. 2.24. *Muon reconstruction efficiency map determined from 2011 data at 7 TeV c.m. energy as a function of muon charge-signed pseudo-rapidity and muon p_T.*

The resultant 2D muon reconstruction efficiency map for 2011 data at 7 TeV c.m. energy is shown in figure 2.24. Significant structure is observed as a function of charge-signed pseudo-rapidity at low p_T and in some limited areas efficiency drops can be seen in the transition region between the barrel and end-cap, the $\eta = 0$ crack region and at the very edge of geometrical acceptance near $|\eta| \sim 2.5$ over the full p_T region studied. Charge-dependent effects persist up to $p_T \sim 11$ GeV beyond which efficiencies are largely symmetric in both the positive and negative-signed pseudo-rapidities. As a function of p_T there are noticeable drops in efficiency at high p_T in the transition region and at the extreme edges of pseudo-rapidity acceptance (albeit with large uncertainties of order 20 % in these regions).

The efficiency to find 4 GeV single muon RoIs is derived using $J/\psi \to \mu^+\mu^-$ decays

fitted from di-muon candidates with $m_{\mu\mu}$ of 2.6 - 4.1 GeV (excluding the ψ' range, 3.5 - 3.75 GeV) in the `EF_mu18` and `EF_mu18 && EF_2mu4_DiMu` trigger samples [90].

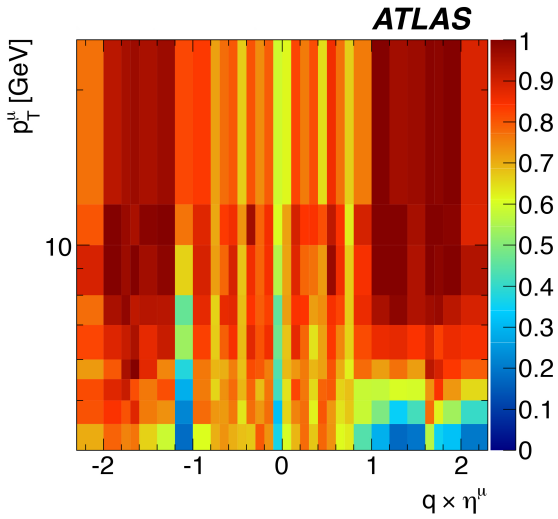


Fig. 2.25. *Single muon RoI efficiency component of* ***EF_2mu4_DiMu*** *di-muon efficiency relative to offline reconstructed muons as a function of single muon charge-signed pseudo-rapidity and transverse momentum measured using 2011 data at 7 TeV c.m. energy [90].*

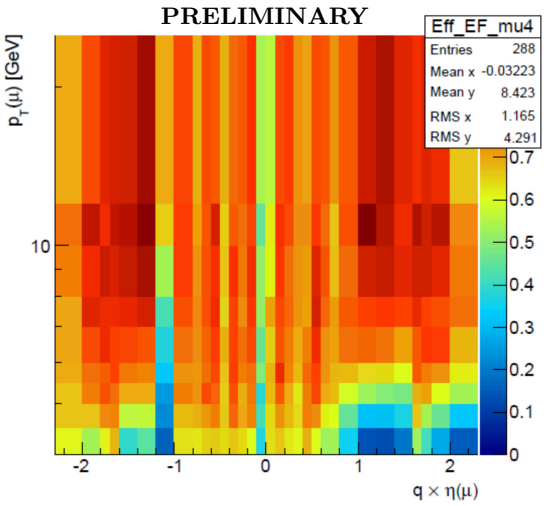


Fig. 2.26. *Single muon RoI efficiency component of* ***EF_2mu4_DiMu*** *di-muon efficiency relative to offline reconstructed muons as a function of single muon charge-signed pseudo-rapidity and transverse momentum measured using 2012 data at 8 TeV c.m. energy (preliminary results).*

Figure 2.25 shows the single muon RoI efficiency component of `EF_2mu4_DiMu` di-muon efficiency relative to offline reconstructed muons as a function of single muon

charge-signed pseudo-rapidity and transverse momentum measured using 2011 data at 7 TeV c.m. energy. Similar charge-dependent pseudo-rapidity structures and drops in efficiency are seen at low p_T in these single muon efficiency maps and regions of localised efficiency loss can be seen in the barrel-endcap transition region near $|\eta| \sim 1.1$ and the crack at $|\eta| \sim 0$. Uncertainties on the values of efficiency in each bin are dominated by the statistics of the `EF_mu18 && EF_2mu4_DiMu` sample.

The efficiency map determined from 2012 data at 8 TeV c.m. energy is shown in figure 2.26, but the analysis is still in progress and the results have not been finalised yet.

2.3.3 Integrated Luminosity in 2011 and 2012 Data Taking Period

An accurate estimation of the integrated luminosity is fundamental for many measurements in ATLAS, like, for example, the cross-sections. Two sets of detectors are installed in ATLAS to measure the instantaneous luminosity delivered by the LHC [87]. The first set uses either event or hit counting algorithms to measure the luminosity on a bunch-by-bunch basis, while the second set infers the total luminosity (summed over all bunches) by monitoring detector currents sensitive to average particle rates over longer time scales. In each case, the detector descriptions are arranged in order of increasing magnitude of pseudo-rapidity.

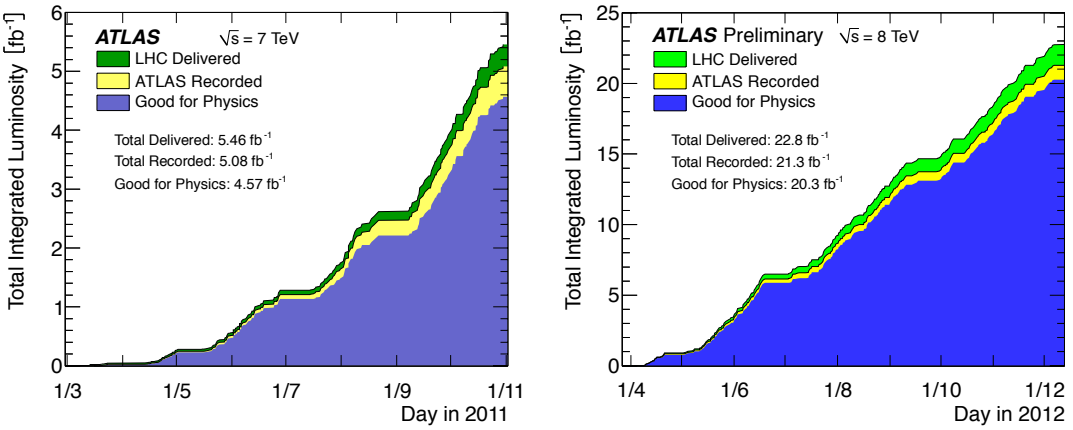


Fig. 2.27. *Cumulative luminosity versus time delivered to (green), recorded by ATLAS (yellow), and certified to be good quality data (blue) during stable beams and for pp collisions at $\sqrt{s} = 7$ TeV in 2011 (left) [87] and at $\sqrt{s} = 8$ TeV in 2012 (right) [86]. The luminosity shown represents the preliminary luminosity calibration. Data quality has been assessed after reprocessing.*

Figure 2.27 shows the integrated luminosity collected in 2011 (left) and 2012 (right). The *delivered luminosity* reported in the plots is the luminosity delivered from the start of the stable beam until the beam dump, when the detector goes into a safe standby mode. The *recorded luminosity* reflects the Data AcQuisition system (DAQ) and the

"warm-start" [15] inefficiencies. The *good for physics* luminosity is the one obtained after having applied data quality criteria to all reconstructed physics objects (an analysis based on a subset of physics objects may use a slightly larger integrated luminosity).

[15] When the stable beam flag is raised, the tracking detectors undergo a so-called "warm-start", which includes a ramp of the high-voltage and, for the pixel, turning on the preamplifiers.

Chapter 3

Lorentz Angle Measurements for the ATLAS SCT

The presence of a magnetic field in silicon detectors influences electrons and holes in the bulk, making them move at a non negligible angle with respect to the electric field direction (see figure 3.1). This angle ϕ_L is called Lorentz angle and influences the properties of the detector such as cluster[1] size and spatial resolution. The knowledge of this quantity is therefore necessary to simulate the detector response, understand the alignment, and perform precise track measurements.

The Lorentz angle has been evaluated analysing SCT cluster widths as a function of the incident angle of the track data collected during 2011 and 2012 LHC pp collisions. When the incident angle equals the Lorentz angle, all charge carriers generated by the particle drift along the particle direction and, apart from charge diffusion, are collected at the same point on the sensor surface, giving a minimum cluster size. The tilt of the barrel modules with respect to the radial direction and correlations between particle transverse momentum and the incident angle for particles originating near the beam axis give rise to a range of possible incident angles for positive and negative particles. Reconstructed tracks are required to have $p_T > 400$ MeV, limiting the possible range of incident angles to approximately $-30° < \phi < -5°$ for positive particles and $-15° < \phi < 10°$ for negative particles, where ϕ is the particle incident angle defined in the module local coordinate system (see section 3.1). Only tracks of negatively charged particles have a minimum in the cluster width distribution and therefore can be used to measure the Lorentz angle in collision data. The minimum of the cluster width distributions is extracted using a binned fit (see section 3.3). The measured values have been compared with two different theoretical models that provide the mobility as a function of the silicon properties.

Section 3.1 gives more details about the effect that the magnetic field produces on the tracks. Section 3.2 analyses the mobility models that have been used to evaluate the expected values of the Lorentz angle, with particular attention to the Becker-

[1] A cluster is a group of contiguous SCT strips with hits, with the cluster width (or size) being the number of strips within it.

Fretwurst-Klanner model that describes the behaviour of sensors with <111> and <100> crystal orientations, and that, for the first time, was adopted in these studies. Section 3.3 describes the procedure used to measure the Lorentz angle and the results obtained with 2011 and 2012 data. Finally, in section 3.4, a comparison with the available CMS measurements is reported.

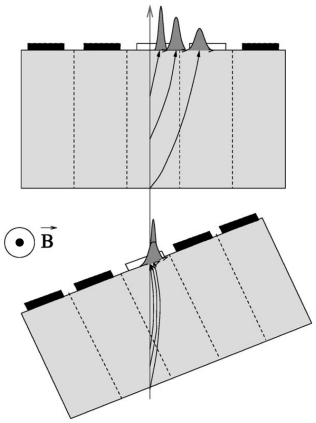


Fig. 3.1. *Representation of the effect of a magnetic field on charge carrier drift. Upper figure shows the spread of carriers (collected by two adjacent strips), generated by a track that crosses the sensor perpendicularly. In the lower, the incident particle crosses the sensor at the Lorentz angle [91].*

3.1 The Lorentz Angle in the Silicon Trackers

The Lorentz angle measurement is particularly important for barrel modules where the direction of the charge carrier drift is perpendicular to the magnetic field generated by the central solenoid. For end-cap modules, charge carriers drift in a direction almost parallel to the magnetic field, thus no significant effect of the magnetic field is expected.

The Lorentz angle for barrel SCT modules is estimated to be about 4 degrees, which corresponds to a 20 μm shift at maximum over the 80 μm strip pitch and the 285 μm sensor width. The Lorentz angle depends on the detector operating conditions, such as bias voltage and temperature, through the electron and hole mobilities. Of course, it also depends on the external magnetic field.

The mean temperature of the silicon sensors in the first three barrel layers is $\sim 270\,\mathrm{K}$ [2], while a higher temperature is maintained in the outermost one ($\sim 278\,\mathrm{K}$) [3], resulting in a lower expected Lorentz angle.

In the local coordinates, the y-axis corresponds to the strip direction, the x-axis is always in the modules plane and perpendicular to the y-axis, and the z-axis is thus parallel to the normal to the sensor plane by construction.

[2] This temperature was chosen to moderate the effects of radiation damage.

[3] The last layer of strips is operated at a higher temperature to act as a thermal shield between the SCT and the TRT which has to operate at room temperature.

3.2 Lorentz Angle Models

The Lorentz angle ϕ_L can be expressed as a function of the mobility of charge carriers as well as the external magnetic field B (in Gauss) as

$$\tan \phi_L = \mu_H \cdot B = r \cdot \mu_d \cdot B \quad , \tag{3.1}$$

where μ_H is the Hall mobility, μ_d (in $cm^2/(V \cdot s)$) the charge carrier mobility and r represents the Hall scattering factor which is around 0.72 for holes [4]. The drift mobility

$$\mu_d = \frac{\mu^0}{\left(1 + \left(\frac{\mu^0 \cdot E}{v^{sat}}\right)^{\beta}\right)^{1/\beta}} \tag{3.2}$$

depends on the low field mobility μ^0, the electric field in the sensor E, the saturation velocity v^{sat} and the β parameter that expresses the transition between ohmic behaviour and saturation velocity [92].

The expected values of the drift mobility for the SCT have been evaluated using two different models: the *Jacoboni-Canali* parameterisation (see section 3.2.1) that allows to describe the behaviour of the modules with crystal orientation <111>, and the *Becker-Fretwurst-Klanner* model that contains the description for both <111> and <100> silicon modules and that will be detailed in section 3.2.2.

Calculations of electric field in strip detectors show that the field is close to what is expected for a flat diode in most of the volume except the region close to the strip. The electric field along the direction z, perpendicular to the sensor plane (see figure 3.1 (a)), is given by [92]:

$$E_z = \frac{V_B + V_D}{d} - 2\,\frac{V_D}{d^2} \cdot z \quad , \tag{3.3}$$

where V_D is the depletion voltage (70 V), V_B is the bias voltage with its nominal value of 150 V, and d is the sensor thickness (285 μm). The expression reported in equation 3.3 is an approximation for highly segmented detectors because of the perturbation of the field near the highly doped collection zones of a thickness of few μm (a few percent of the sensor thickness). All sensors are fully depleted and the Lorentz angle is estimated considering the mean value of the electric field in the sensor:

$$< E_z > = \frac{1}{d} \int_0^d dz\, E_z = \frac{V_B}{d} = 5263 \text{ [V/cm]} \quad . \tag{3.4}$$

[4] The Hall scattering factor depends on the scattering cross section of charge carriers and can be expressed as $r = <\tau^2> / <\tau>^2$ where τ is the mean free time between collisions. It can be evaluated using the dependence of τ on charge carriers energy and their energy distribution.

3.2.1 The Jacoboni-Canali Model

The Jacoboni-Canali model [93, 94] for holes is based on the follow parameterisation as a function of the temperature T:

$$v^{sat}\,(cm/s) = 1.62 \cdot 10^8 \cdot T^{-0.52} \quad (3.5)$$

$$E^c\,(V/cm) = 1.24 \cdot T^{1.68} \quad (3.6)$$

$$\beta = 0.46 \cdot T^{0.17} \quad (3.7)$$

$$\mu_0 = \frac{v^{sat}}{E^c} \quad , \quad (3.8)$$

where E^c is the electric field scale that expresses the transition between the region with a linear and a saturated dependence of the drift velocity on the electric field.

The expected values for the mobility and the Lorentz angle are reported in table 3.1. The drift velocity uncertainty dominates the systematic errors, along with the non uniformity of the electric field E both set to 5 %[5]. The temperature T and the magnetic field B also contribute to systematic error in the Lorentz angle, but in smaller proportion, respectively 0.5 %[6] and 0.75 % [59]. The uncertainties on the bias voltage and impurity in concentration are negligible.

Barrel	**T [K]**	μ_d **[cm²/(V· s)]**	ϕ_L **[deg]**
3	270.9	471 ± 36	3.88 ± 0.29
4	271.3	470 ± 35	3.87 ± 0.29
5	271.7	469 ± 35	3.86 ± 0.29
6	278.6	448 ± 32	3.69 ± 0.26

Table 3.1. *Expected drift mobility and Lorentz angle for the Jacoboni-Canali model.*

3.2.2 The Becker-Fretwurst-Klanner Model

Using the Becker-Fretwurst-Klanner model [95], the drift mobility is estimated starting from the parameters measured at 300 K and summarised in table 3.2. The expected values of the parameters for a certain SCT layer temperature are evaluated using the exponential law:

$$\text{Parameter}_i\,(\text{T}) = \text{Parameter}_i\,(300\ \text{K}) \cdot \left(\frac{T}{300}\right)^{\alpha_i} \quad (3.9)$$

[5]The systematic uncertainty on the drift velocity has been evaluated analysing five different articles (following the reference given by Jacoboni [93]) in which this quantity was studied using two different techniques, one based on Time of Flight (ToF) measurements and the other on conductivity measurements. The former gives the lower systematic bound equal to 5 % while the latter, due to the less accuracy in the determination of the drift velocity, gives the upper bound of around 12 %. In the determination of the uncertainty for the drift velocity, only the ToF method was taken into account and thus the value of 5 % is taken, which is already very conservative.

[6]The mean temperature is measured by sensors mounted on the hybrid of each module [77].

where the values of the exponents α_i, that depend on the parameter under study ($i = \mu_0, v^{sat}, \beta^{<100>}, \beta^{<111>}$), are reported in table 3.3.

	μ_0 [cm^2/(V · s)]	v^{sat} [cm/s]	$\beta^{<100>}$	$\beta^{<111>}$
Holes	474 ± 10	$(0.940 \pm 0.027) \cdot 10^7$	1.181 ± 0.003	0.924 ± 0.002

Table 3.2. *Measured parameters at 300 K derived from the time of flight measurements of drifting charge carriers in planar p^+nn^+ diodes [95].*

	α_{μ_0}	$\alpha_{v^{sat}}$	$\alpha_{\beta^{<100>}}$	$\alpha_{\beta^{<111>}}$
Holes	-2.619 ± 0.007	-0.226 ± 0.002	0.644 ± 0.003	0.550 ± 0.002

Table 3.3. *Values of the α_i parameters derived from the time of flight measurements of drifting charge carriers in planar p^+nn^+ diodes in the temperature range between $-30°$ C and $50°$ C. [95].*

Taking the same systematic errors used to evaluate the uncertainty for the Jacoboni-Canali model (7 %), the expected values for the mobility and the Lorentz angle are summarised in table 3.4.

Barrel	T [K]	<111>		<100>	
		μ_d [cm^2/(V· s)]	ϕ_L [deg]	μ_d [cm^2/(V· s)]	ϕ_L [deg]
3	270.9	425 ± 41	3.50 ± 0.34	488 ± 39	4.01 ± 0.33
4	271.3	424 ± 41	3.50 ± 0.34	487 ± 39	4.01 ± 0.32
5	271.7	423 ± 40	3.49 ± 0.33	485 ± 39	4.00 ± 0.32
6	278.6	407 ± 37	3.35 ± 0.30	464 ± 34	3.82 ± 0.28

Table 3.4. *Expected drift mobility and Lorentz angle for the Becker-Fretwurst-Klanner model.*

3.3 Lorentz Angle Measurements in ATLAS

The detector information used to extract the Lorentz angle is the SCT cluster width. In the absence of a magnetic field the minimum cluster width occurs at an incident angle of zero degrees, where the incident angle is measured in the plane defined by the normal to the wafer surface and the axis (in the wafer plane) perpendicular to the strip direction. The presence of electric and magnetic fields generates a Lorentz force in the silicon strips, therefore the minimum cluster width no longer occurs at zero incident angle, but at the value of the Lorentz angle.

The value of the Lorentz angle has been extracted fitting the distribution of the average cluster width as a function of the incident angle using an empirical function [96]:

$$\begin{aligned} f(\phi) &= (a\,|\tan\phi - \tan\phi_L\,| + b) \otimes \text{Gaussian}(\phi) \\ &= \int_{-\infty}^{\infty} (a\,|\tan\phi' - \tan\phi_L\,| + b) \cdot \frac{1}{\sqrt{2\pi}\,\sigma} \exp\left(-\frac{(\phi-\phi')^2}{2\sigma^2} \right) d\phi' \quad , \end{aligned} \tag{3.10}$$

where ϕ_L is the Lorentz angle, a represents a slope parameter, b is the minimum cluster width and σ is a smearing term that takes into account effects such as the charge carrier diffusion. These four parameters are free and obtained from data. The cluster width distributions are created using the standard ROOT TProfile [97]. Figure 3.2 reports the results of the fit obtained on the four SCT layers, considering the information of the inner side of the modules [7], for sensors with <111> and <100> crystal orientation respectively. The numbering scheme for barrel layers corresponds to the one reported in figure 2.15.

The fit range ($-9° < \phi < +2°$), adopted for all these studies, has been optimised performing a scan on the lower (from $-10°$ and $-6°$) and on the upper (from $-2°$ and $+8°$) bound, and looking at the fit parameters and the χ^2 as a function of the considered fit range. Due to the large statistic available, the fit showed to be extremely robust and stable, with only up to 2 % variation of the Lorentz angle value in the ranges considered for the optimisation.

The stability of the results has also been tested varying the width of the bins in the profile plot (the default used for these studies is 0.5°), and checking the effect of the δ-rays on the fit performance. More details are reported respectively in appendices A.3 and A.4.

The systematic uncertainties related to the choice of the fit model have been estimated performing the fit with an asymmetric function in which two different slope parameters a are considered, below and above the minimum cluster width (more details are provided in section 3.3.1). The mean difference in measured Lorentz angle in each barrel layer is taken as a systematic uncertainty *correlated* among the data taking periods, while the RMS spread of the differences is used as an estimate of the *un-correlated* systematic associated with each dataset. An additional source of correlated systematic among all measurements has been evaluated analysing cosmic ray runs where the magnetic field was off. In these conditions, the charge carriers are not deflected, thus the minimum of the cluster width must occur at null incident angle. Combining all SCT layers, the Lorentz angle extracted from the fit of these data is 0.07 ± 0.05, giving a systematic upper bound of around 0.12 degrees that is added to the correlated systematic evaluated using the asymmetric function.

The plots in figure 3.3 show the behaviour of the Lorentz angle during 2011 and 2012 ATLAS data taking period for the four SCT layers. Within the errors, the

[7] For each module, the side closer to the beam axis is named "inner side", while the other one "outer side". Since the sensors in the inner an outer side are identical, no differences are expected in the Lorentz angle measurements. Nonetheless, the fit has been performed separately on each side of the modules.

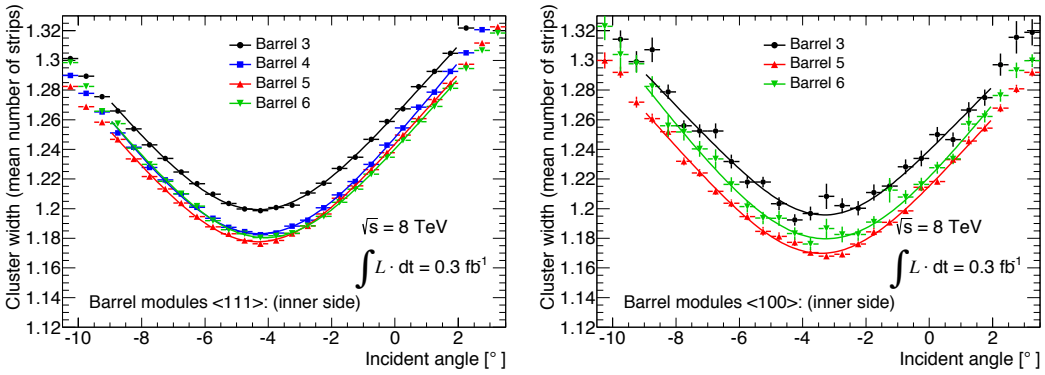


Fig. 3.2. *Lorentz angle fit on the four SCT barrel layers (inner side) for modules with <111> (left) and <100> (right) crystal orientation. The displacement of the minimum from zero is the Lorentz angle ϕ_L. Starting from the same integrated luminosity, the statistics available for the <100> fit is much lower due to the very small number (92) of modules <100> present in the SCT barrel.*

Lorentz angle is clearly constant and does not change as the radiation damage increases (appendices A.1 and A.2 report respectively the behaviour and the correlation of the fit parameters during 2011 and 2012).

The measured values (mean of 2011 and 2012 results) are summarised in table 3.5, while the comparison between the measurements and the expectations is reported in figure 3.4.

Barrel	**T [K]**	**<111>**	**<100>**
3	270.9	4.28 ± 0.23	3.25 ± 0.23
4	271.3	4.26 ± 0.22	-
5	272.7	4.30 ± 0.15	3.28 ± 0.15
6	278.6	4.07 ± 0.15	3.24 ± 0.15

Table 3.5. *Measured values of the Lorentz angle in the SCT barrel for <111> and <100> modules. There are no modules with <100> crystal orientation in barrel 4.*

Figure 3.5 compares the measured and the expected values obtained for the modules <111> and <100>. The measurements are compatible with the model predictions within at most twice the estimated uncertainties on these predictions. The measured values of the Lorentz angle for the <100> sensors are approximately 1° lower than those obtained for <111> sensors. This is contrary to the expectation of the Becker-Fretwurst-Klanner model which states that to a higher expected charge-carrier mobility in the <100> sensors should correspond a higher value of the Lorentz angle.

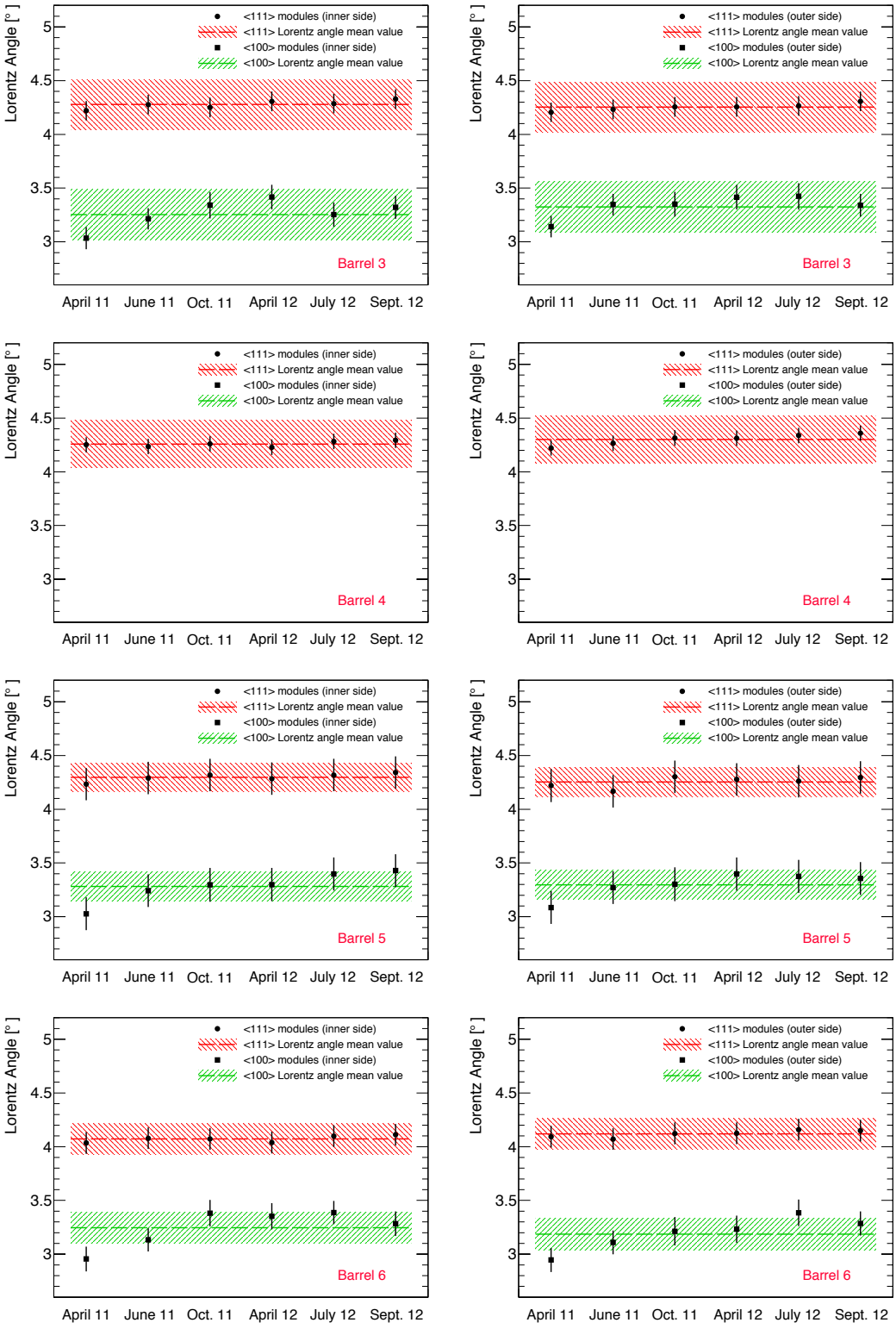


Fig. 3.3. *Lorentz angle measured values in six different runs during 2011/2012 data taking period for the four SCT barrel layers. Plots on the left are for the inner side of the modules whereas plots on the right are for outer side. The error bars are the combination of statistical and un-correlated systematic uncertainties while the dashed lines represent the mean value of the Lorentz angle in 2011 and 2012 with the related correlated systematic (shaded bands). The Lorentz angle does not depend on the increasing radiation damage.*

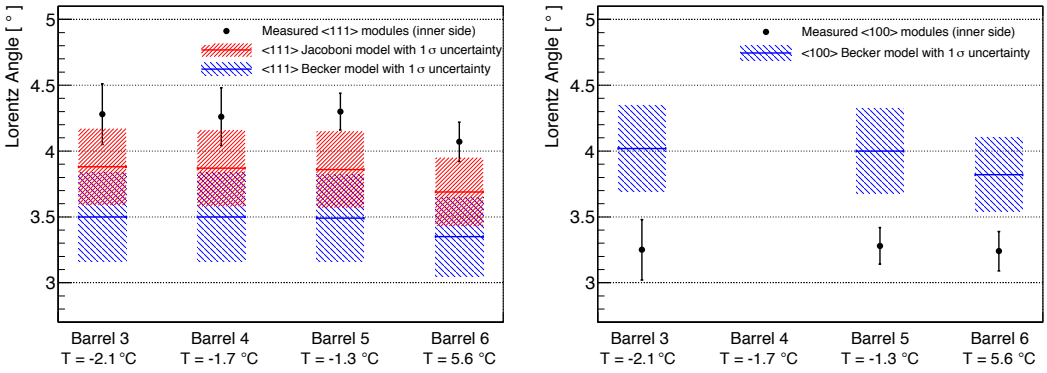


Fig. 3.4. *Comparison between the measured values of the Lorentz angle and the expected ones obtained with the two models analysed in sections 3.2.1 and 3.2.2. The plots show the results for modules <111> (left) and <100> (right). The red and blue lines show the expectations of the Jacoboni-Canali and Becker-Fretwurst-Klanner models respectively, with the* $1\,\sigma$ *uncertainties indicated by the hashed areas. The mean temperature of each layer is shown with the barrel number; a higher temperature is maintained in the outermost layer, barrel 6, resulting in a lower expected Lorentz angle. The error bars of the points include the statistical and systematic uncertainties.*

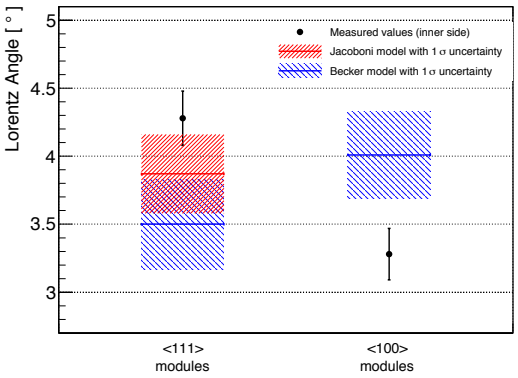


Fig. 3.5. *Comparison between <111> and <100> modules for the measured and the expected values. The plotted values are the mean of the first three SCT barrel layers (for simplicity, the fourth layer has not been included in this plot). The red and blue lines show the expectations of the Jacoboni-Canali and Becker-Fretwurst-Klanner models respectively, with the* $1\,\sigma$ *uncertainties indicated by the hashed areas. The error bars of the points include the statistical and systematic uncertainties.*

3.3.1 Studies on the Fit Systematic

In the past analyses, the only systematic associated with the Lorentz angle measurements was the one evaluated using the cosmic ray runs with the magnetic field off. Since in these conditions the expected Lorentz angle is zero, the deviation of the value obtained by the fit allows to quantify the intrinsic bias of the whole fit procedure. Nevertheless, for a complete evaluation, it was decided to investigate, in more details,

the other possible sources of systematic, in order to be able to give an estimate of both the correlated and the un-correlated uncertainties.

Since the distribution of the cluster width as a function of the incident angle ϕ is clearly asymmetric with respect to the minimum of the cluster width (see figure 3.2), it is interesting to quantify which is the systematic error associated with the symmetric function currently used in the analysis to extract the Lorentz angle. For that reason, the fit of the distributions has been performed using an asymmetric function

$$f(\phi) = \begin{cases} (a_1 \cdot |\tan\phi - \tan\phi_L| + b) \otimes \text{Gauss}(\phi) & (\phi < \phi_L) \\ (a_2 \cdot |\tan\phi - \tan\phi_L| + b) \otimes \text{Gauss}(\phi) & (\phi \geqslant \phi_L) \end{cases} , \tag{3.11}$$

in which the slope a_i $(i = 1, 2)$ is able to float independently below and above the minimum of the cluster width.

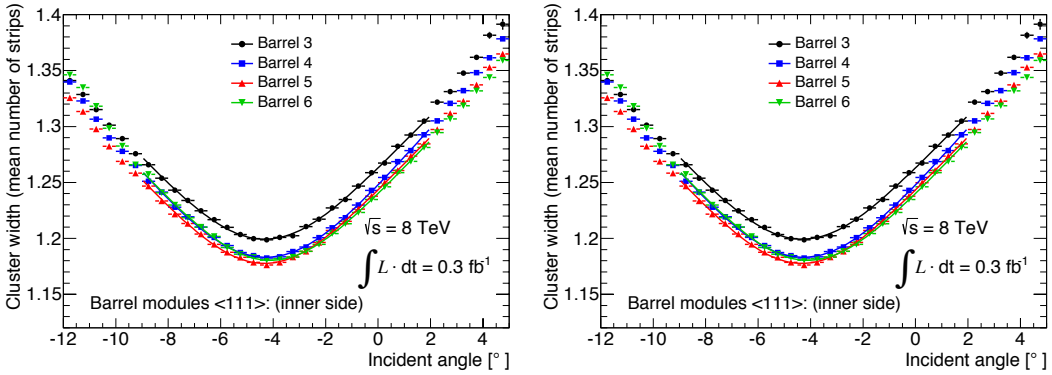


Fig. 3.6. *Lorentz angle fit performed using the standard symmetric function (left) and the alternative asymmetric function (right).*

BA	Symmetric θ_L	Symmetric χ^2/NDF	Asymmetric θ_L	Asymmetric χ^2/NDF	Δ [%]
3	- 4.33 ± 0.01	1.37	- 4.15 ± 0.06	0.944	-0.18 [4.2]
4	- 4.29 ± 0.01	1.59	- 4.05 ± 0.06	0.852	-0.24 [5.6]
5	- 4.34 ± 0.01	1.11	- 4.22 ± 0.02	0.929	-0.12 [2.8]
6	- 4.11 ± 0.01	0.93	- 4.15 ± 0.06	1.57	-0.04 [1.0]

Table 3.6. *Lorentz angle values, for <111> modules, obtained with the standard and the asymmetric function for the runs 210302 and 210308 collected in September 2012. The χ^2/NDF of the fits and the differences (with sign and in %) between the two results are also reported.*

Figure 3.6 shows the comparison of the fit performed using the standard and the asymmetric fit function on data collected in September 2012 (run 210302 + run 210308)[8]. The results obtained with the two fits are summarised in table 3.6. The

[8]This is the same dataset used in the plots in figure 3.2.

analysis of the χ^2/NDF (close to 1 for both procedures) shows how the performance of the two strategies are comparable. Nevertheless, the error of the asymmetric fit is higher than the standard one due to the additional parameter introduced to model the two different slopes.

		Inner side			Outer side		
	BA	**Symmetric**	**Asymmetric**	**Δ [%]**	**Symmetric**	**Asymmetric**	**Δ [%]**
April 2011	3	- 4.22 ± 0.01	- 3.98 ± 0.04	- 0.24 [5.7]	- 4.21 ± 0.01	- 4.06 ± 0.04	- 0.15 [3.6]
	4	- 4.25 ± 0.01	- 4.09 ± 0.02	- 0.16 [3.8]	- 4.22 ± 0.01	- 4.10 ± 0.04	- 0.12 [2.8]
	5	- 4.23 ± 0.01	- 4.28 ± 0.06	0.05 [1.2]	- 4.22 ± 0.01	- 4.27 ± 0.05	0.05 [1.2]
	6	- 4.03 ± 0.01	- 3.94 ± 0.06	- 0.09 [2.2]	- 4.09 ± 0.01	- 4.26 ± 0.01	0.17 [4.2]
June 2011	3	- 4.28 ± 0.01	- 3.99 ± 0.04	- 0.29 [6.8]	- 4.23 ± 0.01	- 4.16 ± 0.06	- 0.07 [1.7]
	4	- 4.24 ± 0.01	- 4.16 ± 0.02	- 0.08 [1.9]	- 4.27 ± 0.01	- 4.10 ± 0.04	- 0.17 [4.0]
	5	- 4.29 ± 0.01	- 4.23 ± 0.04	- 0.06 [1.4]	- 4.17 ± 0.02	- 4.55 ± 0.04	0.38 [9.1]
	6	- 4.08 ± 0.01	- 3.91 ± 0.04	- 0.17 [4.2]	- 4.07 ± 0.01	- 3.98 ± 0.05	- 0.09 [2.2]
Oct. 2011	3	- 4.25 ± 0.02	- 3.96 ± 0.08	- 0.29 [6.8]	- 4.26 ± 0.02	- 4.25 ± 0.10	- 0.01 [0.2]
	4	- 4.26 ± 0.02	- 4.07 ± 0.08	- 0.19 [4.5]	- 4.31 ± 0.01	- 4.14 ± 0.09	- 0.17 [3.9]
	5	- 4.32 ± 0.02	- 4.25 ± 0.08	- 0.07 [1.6]	- 4.30 ± 0.02	- 4.36 ± 0.06	0.06 [1.4]
	6	- 4.07 ± 0.02	- 3.98 ± 0.08	- 0.09 [2.2]	- 4.12 ± 0.02	- 4.05 ± 0.10	- 0.07 [1.7]
April 2012	3	- 4.31 ± 0.02	- 3.96 ± 0.09	- 0.35 [8.1]	- 4.26 ± 0.02	- 4.08 ± 0.09	- 0.18 [4.2]
	4	- 4.23 ± 0.01	- 4.08 ± 0.07	- 0.15 [3.6]	- 4.31 ± 0.02	- 4.18 ± 0.08	- 0.13 [3.0]
	5	- 4.28 ± 0.02	- 4.14 ± 0.07	- 0.14 [3.3]	- 4.28 ± 0.02	- 4.46 ± 0.05	0.18 [4.2]
	6	- 4.04 ± 0.01	- 3.85 ± 0.07	- 0.19 [4.7]	- 4.12 ± 0.02	- 4.05 ± 0.08	- 0.07 [1.7]
July 2012	3	- 4.29 ± 0.01	- 4.11 ± 0.07	- 0.18 [4.2]	- 4.27 ± 0.02	- 4.06 ± 0.06	- 0.21 [4.9]
	4	- 4.28 ± 0.01	- 3.91 ± 0.05	- 0.37 [8.6]	- 4.34 ± 0.01	- 4.12 ± 0.07	- 0.22 [5.1]
	5	- 4.32 ± 0.01	- 4.12 ± 0.05	- 0.20 [4.6]	- 4.26 ± 0.02	- 4.27 ± 0.07	0.01 [0.2]
	6	- 4.10 ± 0.01	- 3.91 ± 0.07	- 0.19 [4.6]	- 4.16 ± 0.02	- 4.09 ± 0.08	- 0.07 [1.7]
Sept. 2012	3	- 4.33 ± 0.01	- 4.15 ± 0.06	- 0.18 [4.2]	- 4.31 ± 0.01	- 4.03 ± 0.07	- 0.28 [6.5]
	4	- 4.29 ± 0.01	- 4.05 ± 0.06	- 0.24 [5.6]	- 4.36 ± 0.01	- 4.18 ± 0.03	- 0.18 [4.1]
	5	- 4.34 ± 0.01	- 4.22 ± 0.02	- 0.12 [2.8]	- 4.30 ± 0.01	- 4.24 ± 0.01	- 0.06 [1.4]
	6	- 4.11 ± 0.01	- 4.15 ± 0.06	- 0.04 [1.0]	- 4.15 ± 0.01	- 3.96 ± 0.06	- 0.19 [4.6]

Table 3.7. *Lorentz angle values, for <111> modules and each barrel layer (BA), obtained with the standard and the asymmetric function for the different data-taking periods analysed. The differences (with sign and in %) between the two results are also reported.*

The same studies have been performed also for all the other SCT layers and sides and for all data taking periods considered in this analysis. The results for modules <111> and <100> are summarised in table 3.7 and 3.8 respectively, that also report the deviation of the Lorentz angle extrapolated with the asymmetric fit with respect to the value obtained with the standard function. These results show how the differences

		Inner side			Outer side		
	BA	**Symmetric**	**Asymmetric**	**Δ [%]**	**Symmetric**	**Asymmetric**	**Δ [%]**
April 2011	3	- 3.03 ± 0.05	- 3.24 ± 0.17	0.21 [6.9]	- 3.14 ± 0.04	- 3.33 ± 0.31	0.19 [6.1]
	4	-	-	-	-	-	-
	5	- 3.03 ± 0.02	- 3.21 ± 0.04	0.18 [5.9]	- 3.09 ± 0.02	- 3.33 ± 0.12	0.24 [7.8]
	6	- 2.96 ± 0.05	- 3.39 ± 0.30	0.43 [14.5]	- 2.95 ± 0.05	- 2.67 ± 0.07	- 0.28 [9.5]
June 2011	3	- 3.21 ± 0.05	- 3.26 ± 0.13	0.05 [1.6]	- 3.35 ± 0.04	- 3.38 ± 0.19	0.03 [0.9]
	4	-	-	-	-	-	-
	5	- 3.24 ± 0.02	- 3.32 ± 0.10	0.08 [2.5]	- 3.27 ± 0.02	- 3.07 ± 0.12	- 0.20 [6.1]
	6	- 3.14 ± 0.04	- 2.73 ± 0.04	- 0.41 [13.1]	- 3.11 ± 0.05	- 2.88 ± 0.14	- 0.23 [7.4]
Oct. 2011	3	- 3.34 ± 0.08	- 3.26 ± 0.25	- 0.08 [2.4]	- 3.35 ± 0.05	- 3.41 ± 0.23	0.06 [1.8]
	4	-	-	-	-	-	-
	5	- 3.30 ± 0.05	- 3.22 ± 0.06	- 0.08 [2.4]	- 3.30 ± 0.02	- 3.36 ± 0.11	0.06 [1.8]
	6	- 3.38 ± 0.06	- 3.73 ± 0.18	- 0.35 [10.4]	- 3.21 ± 0.11	- 3.39 ± 0.07	0.18 [5.6]
April 2012	3	- 3.42 ± 0.07	- 3.53 ± 0.29	0.11 [3.2]	- 3.42 ± 0.07	- 3.08 ± 0.17	- 0.34 [14.6]
	4	-	-	-	-	-	-
	5	- 3.30 ± 0.06	- 3.46 ± 0.15	0.16 [4.5]	- 3.40 ± 0.04	- 3.02 ± 0.05	- 0.38 [11.2]
	6	- 3.34 ± 0.05	- 3.28 ± 0.31	- 0.06 [1.8]	- 3.33 ± 0.03	- 2.98 ± 0.16	- 0.35 [10.5]
July 2012	3	- 3.25 ± 0.07	- 3.00 ± 0.42	- 0.25 [7.7]	- 3.42 ± 0.08	- 3.27 ± 0.25	- 0.15 [4.4]
	4	-	-	-	-	-	-
	5	- 3.40 ± 0.03	- 3.46 ± 0.16	0.06 [1.8]	- 3.38 ± 0.04	- 3.40 ± 0.11	0.02 [0.6]
	6	- 3.39 ± 0.06	- 3.08 ± 0.10	- 0.31 [9.2]	- 3.38 ± 0.04	- 3.26 ± 0.22	- 0.12 [3.6]
Sept. 2012	3	- 3.32 ± 0.05	- 3.63 ± 0.17	0.31 [9.3]	- 3.34 ± 0.04	- 3.00 ± 0.19	- 0.34 [10.2]
	4	-	-	-	-	-	-
	5	- 3.43 ± 0.03	- 3.34 ± 0.14	- 0.09 [2.6]	- 3.36 ± 0.02	- 3.29 ± 0.02	- 0.07 [2.1]
	6	- 3.28 ± 0.05	- 3.56 ± 0.31	0.28 [8.5]	- 3.29 ± 0.03	- 2.85 ± 0.03	- 0.44 [13.4]

Table 3.8. *Lorentz angle values, for <100> modules and each barrel layer (BA), obtained with the standard and the asymmetric function for the different data-taking periods analysed. The differences (with sign and in %) between the two results are also reported.*

obtained with the asymmetric fit are correlated since, on average, the asymmetric fit provides smaller value of the Lorentz angle. In addition, the asymmetric fit has larger systematic uncertainties that are un-correlated among all periods, and shows to be less stable than the symmetric one. For that reason, a careful estimation of the systematic uncertainty, due to the choice of the fit function, is mandatory.

For each barrel layer, the average in time (among all six data taking periods and both sensor sides) of the difference (reported in table 3.7 and 3.8) between symmetric and asymmetric fit, $<\Delta>$, has been taken as correlated systematic associated with the fit. Since there is no physical reason for the systematic to be different on the two sides

of the sensors, both sides together have been considered in the evaluation of the mean. This also allows to reduce the effect of the statistical fluctuations in the estimation of the correlated systematic. In addition, since for the same integrated luminosity track data for <100> modules are much less than track data for <111> (therefore the Lorentz angle measurement for <100> modules is more sensible to statistical fluctuations) and since, again, there is no physical reason for the systematic error to be different for the <100> modules, the results obtained for <111> has been used also for <100>.

The correlated part of the global systematic has two different components. The first one comes from the fit of the cosmic-ray data with the magnetic field off and the other from the average in time of the differences that we discussed above. Thus the global correlated systematic associated with the Lorentz angle values is the sum in quadrature of the two contributions.

A summary of the correlated systematic for each SCT layer is reported in table 3.9. The red and green bands in the plots reported in figure 3.3 are obtained summing in quadrature the global correlated systematic and the statistical error of the mean value of the Lorentz angle in the six data taking periods.

BA	**Cosmic rays syst.**	$\mathbf{<\Delta>}$	$\boldsymbol{\epsilon}^{\mathbf{correlated}}$
3	0.12	-0.20	0.23
4	0.12	-0.18	0.22
5	0.12	0.01	0.12
6	0.12	-0.08	0.14

Table 3.9. *Summary of the correlated systematic for each SCT barrel layer (BA).*

The un-correlated systematic associated with the fit has been assessed considering the RMS of the shift with respect to the average value <Δ>. This procedure takes into account of the statistical fluctuations that affect the differences between the symmetric and the asymmetric fit. In more detail, for each SCT layer and each data taking period, the average in time difference <Δ> is subtracted from the differences (with sign) between the standard and the asymmetric function Δ (values are summarised in table 3.10), and then divided by two times the number of periods:

$$\epsilon^{\text{un-correlated}} = \sqrt{\frac{1}{12} \cdot \left(\sum_{\text{periods, sides}} (\Delta - <\Delta>)^2 \right)} \quad . \tag{3.12}$$

The final results obtained are summarised in table 3.11.

The global error associated with the Lorentz angle measurement is therefore the sum in quadrature of the statistical error of the fit, the correlated systematic, and the un-correlated systematic.

< 111 > - inner side						
BA	**April 2011**	**June 2011**	**October 2011**	**April 2012**	**July 2012**	**September 2012**
3	- 0.04	- 0.09	- 0.09	- 0.15	0.02	0.02
4	0.02	0.10	- 0.01	0.03	- 0.19	- 0.06
5	0.04	- 0.07	- 0.08	- 0.15	- 0.21	- 0.13
6	- 0.01	- 0.09	- 0.01	- 0.11	- 0.11	0.12
< 111 > - outer side						
3	0.05	0.13	0.19	0.02	- 0.01	- 0.08
4	0.06	0.01	0.01	0.05	- 0.04	0.00
5	0.04	0.37	0.05	0.17	0.00	- 0.07
6	0.25	- 0.01	0.01	0.01	0.01	- 0.11

Table 3.10. *Deviations between the differences standard-asymmetric function reported in tables 3.7 and 3.8 and the average in time of the differences* $<\Delta>$.

BA	$\epsilon^{\text{un-correlated}}$
3	0.09
4	0.07
5	0.15
6	0.10

Table 3.11. *Un-correlated systematic for the four SCT layers.*

3.3.1.1 Other Possible Sources of Systematic Error

Other possible sources of uncertainties, that could be associated with the fit procedure adopted in the analysis, have been investigated.

The Lorentz angle fit has been performed using a parabola to model the cluster width distribution as a function of the incident angle. Unfortunately, this procedure exhibited a systematic difference (that is more or less the same for all layers and sides) between the standard and the parabolic fit that proved the parabola not to be reliable to extrapolate the Lorentz angle (see appendix A.5). Therefore, the results obtained have not been considered in the systematic estimation.

Since the correlated systematic from cosmic ray runs have been evaluated using data collected in early 2011, in order to improve the estimation of this uncertainty it would have been interesting to redo this analysis using 2012 data that take advantage of updated tracking algorithms and detector calibrations. Unfortunately, the small amount of cosmic track data available in 2012 did not allow to perform these studies.

3.4 Comparison of the ATLAS and CMS Results

In order to improve the understanding of the results obtained, the ATLAS and CMS measured values of the Lorentz angle have been compared.

The CMS Silicon Strip Tracker is composed of three sub-detectors: Tracker Inner Barrel (TIB)/Tracker Inner Disk (TID), Tracker Outer Barrel (TOB) and Tracker End Cap (TEC), made of silicon sensors with <100> crystal orientation. In table 3.12 are reported the results [98] for the two barrel detectors, TIB and TOB. The big errors on the CMS expected values are due to a huge systematic uncertainty on the depletion voltage ($\sim 70\,\%$; $V_{\text{depl}} = 150 \pm 100$ V) that leads to a big error in the definition of the electric field in the sensors. Due to the uncertainty in the depletion voltage, some sensors could not be considered as fully depleted, to the contrary of ATLAS, where all sensors are fully depleted. Moreover, CMS quoted an error on the temperature measurement that is bigger than ATLAS (CMS has an error of around 5 % whereas in ATLAS the precision is better than 1 %). The temperature of the CMS modules is around 300 K.

ATLAS p-p collisions (2011 and 2012)			
Barrel	**T [K]**	**<111>**	**<100>**
3	270.9	4.28 ± 0.23	3.25 ± 0.23
4	271.3	4.26 ± 0.22	-
5	271.7	4.30 ± 0.15	3.28 ± 0.15
6	278.6	4.07 ± 0.15	3.24 ± 0.15

CMS cosmic rays (Autumn 2008) [98]			
Barrel	**T [K]**	**<100>**	**Expected**
TIB	298 ± 15	4.0 ± 1.2	5.8 ± 0.6
TOB	298 ± 15	5.1 ± 0.6	6.4 ± 0.6

Table 3.12. *ATLAS and CMS Lorentz angle results.*

Unfortunately, the only public CMS results available at the moment are those obtained with cosmic ray runs collected in 2008 [98], thus a proper comparison cannot be made.

The two plots in figure 3.7 show the results obtained by the two experiments. The plot on the left side contains the measured Lorentz angles that are, as expected, in disagreement due to the different magnetic field in the two experiments (3.7 T in the track detector of CMS and 2 T in ATLAS). For that reason, the tangent of the Lorentz angle divided by the magnetic field (that is the Hall mobility) has been evaluated, and the results are reported in the right plot of figure 3.7. Even after having normalised for the different magnetic field, the angles measured by the two experiments still disagree.

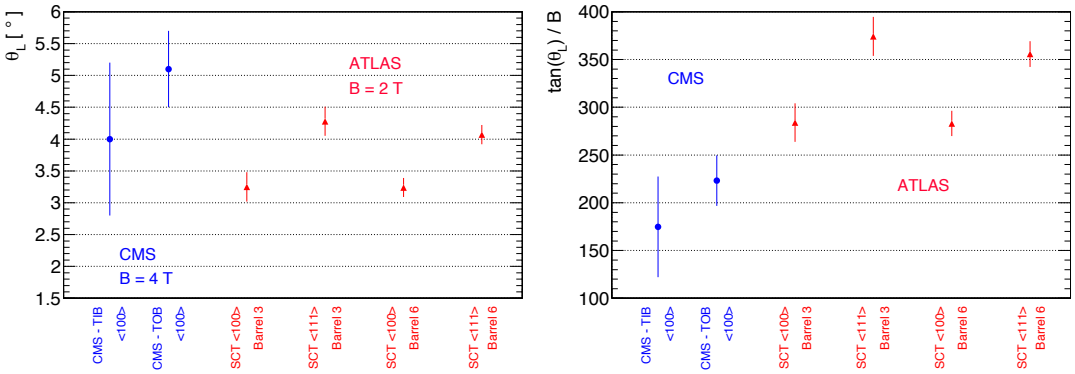


Fig. 3.7. *Comparison of the Lorentz angle measured values (left) and Hall mobility (right) obtained by the ATLAS and CMS experiments.*

3.5 Conclusions

In this chapter I have reported the studies and the measurements on the Lorentz angle that I have performed using LHC 2011 and 2012 collision data collected by ATLAS. The Lorentz angle have been obtained by fitting the distributions of the cluster width as a function of the incident angle for modules with <111> and <100> crystal orientation. The high statistic now available allows to obtain a very precise estimation of this quantity with systematic uncertainties evaluated analysing cosmic ray runs where the magnetic field was off, and studying the differences between the fit performed with the standard and the alternative asymmetric function.

Two different models, used to evaluate the mobility of charge carriers, have been studied. The Jacoboni-Canali parameterisation allows to estimate the Lorentz angle value for the modules with <111> crystal orientation whereas the Becker-Fretwurst-Klanner model (for the first time used in this kind of analysis) is able to describe both <111> and <100> modules. Discrepancies between measured and expected values, obtained with the two models, have been found, but they are within the uncertainty of two standard deviations.

Chapter 4

Backplane Resistance Measurements for the ATLAS Silicon Strips

During the ATLAS assembly in 2006 and 2007, while checking the continuities of cabling and connections by applying a forward voltage[1], some modules exhibited a higher resistance than expected in the bias circuit. In these tests, 10 μA was sourced in the forward bias direction, and the resistance measured. At 10 μA the strip diodes are not fully turned on (which requires $\sim$0.23 V), and the typical resistance is about $\sim$35 k$\,\Omega$ for modules with a negligible resistance of the sensor backplane. However, around 20 % of the barrel modules and 5 % of the end-cap modules showed a much higher resistance value at 10 μA.

In the SCT modules, the bias connection is made through a silver-loaded epoxy to the surface of the aluminised backplane of the silicon sensors that sometimes provoked the high resistance. Fortunately, it was found empirically that, sourcing a large forward bias current or operating the module at higher than normal bias voltage (350 V instead of 150 V) usually reduces the high resistance down to normal levels. However, an exposure to high radiation doses could create higher resistance and, consequently, a drop in the bias voltage applied, which would induce a high leakage current in the modules. For that reason, after Run 1, it was decided to re-check again the resistance of the contacts of a sample of modules.

4.1 Power Supply System

The power supply crate [100] provides independent voltages and control signals necessary to operate 48 SCT modules. The Low Voltage (LV) provides supply voltages

[1]When a silicon sensor (diode) is in *forward bias*, the depletion zone becomes narrower and the current flows from the anode towards the cathode. In *reverse bias* the depletion zone becomes larger and the current is blocked, therefore there is no flow of electricity through the circuit. Silicon sensors are operated in reverse bias mode.

necessary for clock signal, opto-electronics, monitoring of the temperature and the Detector Module ASIC's (Application Specific Integrated Circuit) [101] whereas the High Voltage (HV) provides the bias voltage necessary to completely deplete each silicon sensor. Figure 4.1 describes the layout of a single crate. Each crate is composed of 12 LV cards with four channels each and 6 HV cards with eight channels each.

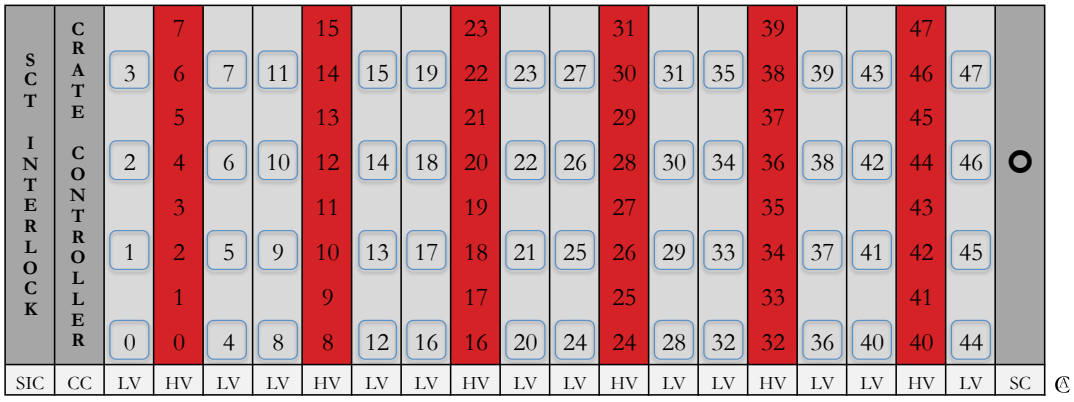


Fig. 4.1. *SCT power supply crate. In red are reported the 6 HV cards whereas in light grey the 12 LV cards.*

4.2 Barrel Modules and Bias Connections

The two pairs of SCT sensors (see also section 2.2.3.2) are glued on a 380 μm-thick Thermal Pyrolytic Graphite (TPG) support that provides the thermal and mechanical structure. A Cu/Polyamide flex-circuit-based hybrid with a carbon-fibre substrate is wrapped around the sensor baseboard sandwich. The two 770-strip (768 active) sensors on each side form a 128 mm long unit (126 mm active with a 2 mm dead space). The high voltage is applied to the sensors via the TPG by means of a silver loaded epoxy adhesive. The figure 4.2 shows the glue process of the baseboard: the four large dots are the silver loaded epoxy to make the electrical contacts with the sensors whereas the small white dots are boron-nitride-doped epoxy adhesives.

Figure 4.3 shows the bias circuit in the hybrid, showing the combined resistance of 11.2 KΩ. In addition, the 770 1.25 MΩ strip resistors in parallel contribute a further $\sim$ 0.4 KΩ.

Due to the diode characteristic, a small forward voltage was applied to the bias circuit throughout the hybrid in order to turn on the diode. The measurements were performed with a small constant current of 10 μA which sometimes proved to be too low to fully turn on the diode. However, relaxing the current limit up to 1 mA most of the modules reduced the resistance toward the expected 12 KΩ.

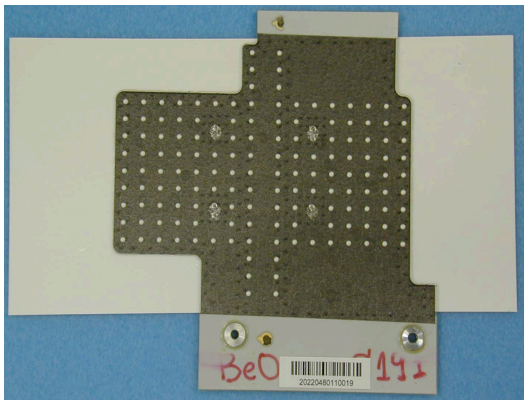


Fig. 4.2. *Glue patterns on the SCT baseboard. The four large dots are silver loaded epoxy to make electrical contacts with the sensor backplane, the white dots are the adhesives [99].*

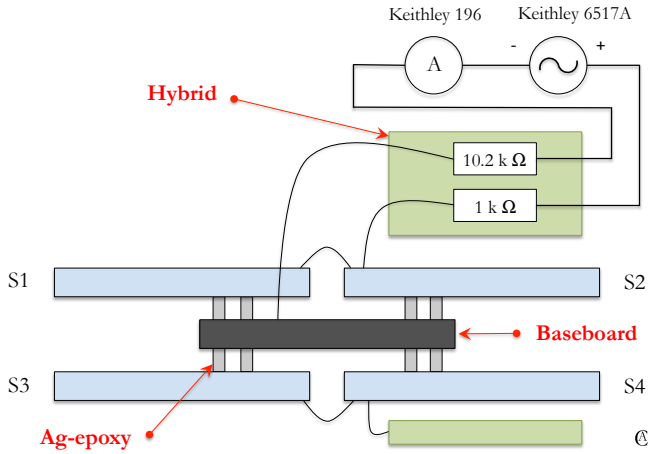


Fig. 4.3. *Setup for continuity measurements. Biasing scheme to sensors (S1, S2, S3 and S4) in SCT barrel modules is illustrated.*

4.3 Backplane Resistance Measurements

The following sections will report the procedure and the apparatus used to measure the backplane resistance of the SCT modules.

4.3.1 Measurement Procedure

The measurements have been performed using exactly the same procedure that was adopted for all 2006 and 2007 tests.

1. First a source current of 10 μA is applied and after 1 s a resistance measurement is performed.

2. If the reading is above 40 kΩ, the current is increased to 100 μA for 3 s.

3. The current is reduced to 10 μA and the measure is repeated.

4. If the measured resistance is still above 40 kΩ, the current is again increased to 100 μA for 10 s.

5. The current is reduced to 10 μA and the last measurement is made.

4.3.2 Measurement Apparatus

A Keithley Unit 2410 [102] is used to measure the HV backplane resistance under forward bias. A Agilent 34970A [103] data acquisition is connected to the pc unit via a National Instrument GPIB-USB-HS [104] that allows the measurement software based on LabVIEW [105] to automatically iterate the procedure reported in section 4.3.1, read the resistance and save all results in a text file that can be subsequently analysed. In order to measure the resistance, each HV board is substituted with another specific board that allows to connect the experimental apparatus directly to the bias circuit (see figure 4.4).

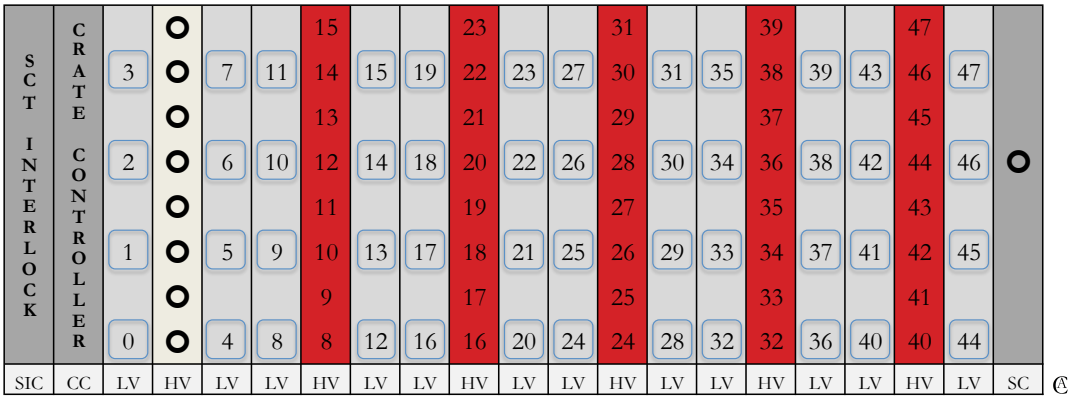


Fig. 4.4. *Example of measurement. The first HV board is substituted with a specific board that allows to directly connect the measurement apparatus to the HV circuit throughout eight connectors (one for each channel).*

4.4 Backplane Resistance Results

The next sections will focus on the measurements performed on the SCT modules installed in the ATLAS cavern USA15 as well as on some test modules present in the SCT clean room at Point 1 (SR1). For each channel, the value of the resistance obtained for each step of the procedure, illustrated in section 4.3.1, is recorded.

4.4.1 Modules Installed in USA15

The results obtained on the barrel and end-cap modules connected to the crates in the ATLAS service cavern (USA15) are shown in the figure 4.5. Both 2006/2007 and

2013 measurements are reported in the plots.

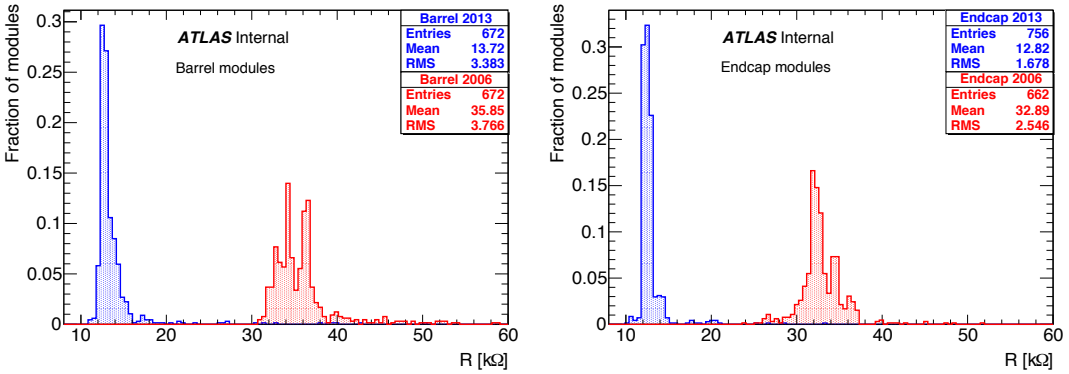


Fig. 4.5. *HV backplane resistance measurements for barrel modules (left) and end-cap modules (right). In blue are reported the current results whereas in red the 2006/2007 measurements.*

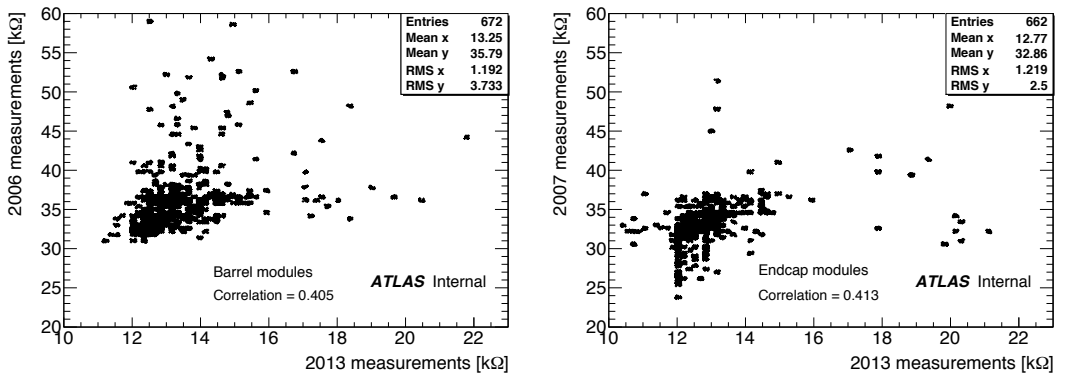


Fig. 4.6. *Scatter plots for barrel (left) and end-cap (right) modules. The correlation between old and new measurements is also reported.*

It can be seen from figure 4.5 that, for those modules with a negligible backplane resistance, the measured resistance values at 10 μA forward bias current dropped from about 32-35 KΩ at installation to about 12 KΩ after 3 years of operation at the LHC. This suggests that the forward bias characteristic has changed following irradiation. However, the plots show that a small number of modules exhibit high backplane resistance in 2013 (corresponding to the outliers of the blue-filled solid histogram in the plot). Figure 4.6 compares measured module resistances at 10 μA forward bias current for 2007 data and 2013 data. The plots suggest that there is no correlation: modules "cured" with a large forward bias current in 2007 do not necessarily show high backplane resistance again, whereas modules which did not originally manifest high resistance can develop it.

The level of radiation background in the ATLAS detector depends on the distance from the interaction point of the two protons. Following the hypothesis of a dependence

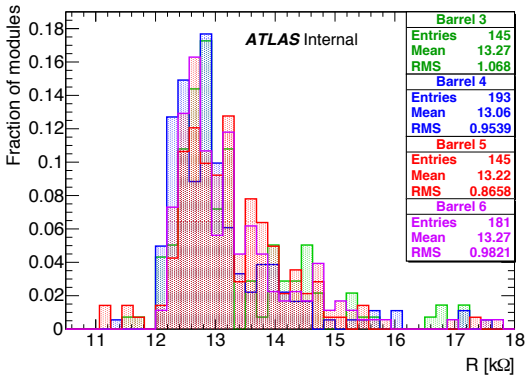


Fig. 4.7. *HV backplane resistance for the four SCT barrel layers.*

of the forward bias characteristic on the ionising dose, the higher is the distance, the higher should be the resistance measured. Figure 4.7 reports the results separated per-layer, and no substantial dependence can be seen.

4.4.2 Modules Installed in SR1

Figure 4.8 shows the results obtained with the modules installed in the overground SCT test area in Point 1, using exactly the same procedure that was adopted for the modules installed in the cavern. Unfortunately, there is no record available of the measurements performed in 2007, but this result shows how the modules that have not been irradiated have a mean value that is fully compatible with the 2006/2007 results, obtained on those modules that were installed in the cavern.

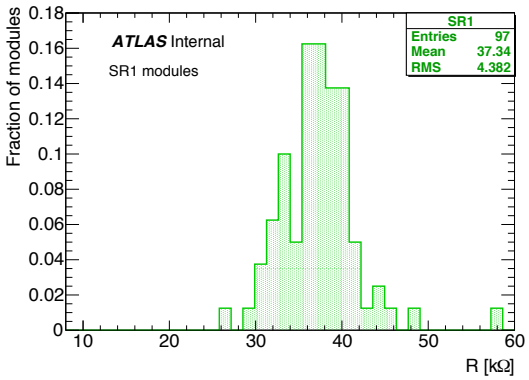


Fig. 4.8. *HV backplane resistance measurements for the modules installed in SR1.*

Figure 4.9 summarises the results of all measurements performed on the modules.

4.4.3 Backplane Resistance of the Cavern Unpowered End-Cap Modules

The modules installed in the cavern have been powered on for a long period of time and the HV and the current could have had an effect on the silver-loaded epoxy reducing the overall resistance. To check this hypothesis, the resistance of some end-cap modules, that have been never powered on due to some cooling problems, have been measured.

Crate	Channels	Mean R [kΩ]	Total mean R [kΩ]
50	$40 \mapsto 46$	12.4	15.0
78	$18 \mapsto 23$	17.5	

Table 4.1. *Summary of the measurements on unpowered modules.*

The measurements performed are summarised in table 4.1. The mean value of the resistance for these modules is around 15 kΩ, very close to the value obtained for the other cavern modules that have been always powered on.

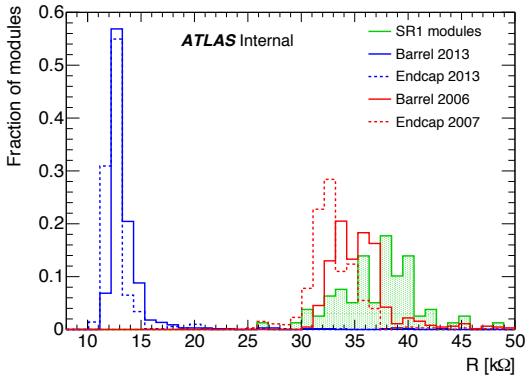


Fig. 4.9. *Summary of all backplane measurements performed on the modules.*

4.4.4 High Resistance Modules Check

Among all cavern modules that have been tested, some showed a resistance higher than 40 kΩ even after having applied a current of 100 μA for 10 s. In order to understand if these modules can be "cured" by a higher current, 1 mA was applied to the bias circuit for few seconds and the measurement at 10 μA was taken again. The results are reported in table 4.2 and show how the higher current can reduce the resistance up to "normal" values.

Crate	Channel	<R> [kΩ] @ 10 μA	<R> [kΩ] @ 10 μA [after 1 mA]
Y.25-11.A2 / 3 (barrel)	45	47	13
Y26-11.A2 / 4 (barrel)	17	42	13
Y29-11.A2 / 3 (barrel)	28	42	13
Y29-14.A2 / 2 (barrel)	10	42	17
Y23-14.A2 / 4 (end-cap)	16	81	17
Y26-14.A2 / 4 (end-cap)	38	50	12

Table 4.2. *Summary of the measurements on the modules with high resistance.*

4.5 Conclusions

Since a high backplane resistance was measured in some modules during the ATLAS assembly in 2006 and 2007, after the first long run of data-taking new measurements have been required to check possible effects of the high radiation levels.

The modules installed in the ATLAS cavern showed a mean resistance of around 12 kΩ, lower than the mean resistance measured in 2006 and 2007 that was around 33-35 kΩ. The differences between the two measurements are probably due to the effect of the high radiation doses received by the modules during the *pp* collisions. This hypothesis was confirmed testing the modules that are installed in SR1 and that have not been irradiated in the past few years. The mean resistance of those is around 37 kΩ, in agreement with the values obtained in the 2006 and 2007 for the modules installed in the ATLAS cavern.

After high radiation doses received during Run 1, the forward bias characteristic appears to have changed for the SCT modules. Therefore, the strip diodes that before irradiation required a higher bias current, after irradiation are fully turned on at 10 μA. A small number of modules still exhibit higher than normal backplane resistance after Run1, but they are not necessarily the same modules that showed high backplane resistance during commissioning in 2007. As before, the high resistance for those modules can be returned to normal levels by applying a higher forward bias current (100 μA or even 1 mA).

The possible effects of the HV and current on the silver loaded contacts have been tested on some end-cap modules that have never been powered on due to cooling problems. The results obtained are in agreement with the mean resistance measured on the other modules.

These new measurements show that, even after the first long period of data taking and thus after very high radiation doses, the backplane resistance does not represent an issue for the future performance of the SCT detector.

Chapter 5

Rare B Decays 2011 Analysis and Combinatorial Background Studies

In this chapter I will review the search for the $B_s^0 \to \mu^+\mu^-$ rare decay performed on 2011 data. Every step of the analysis flow is briefly treated here, but I will focus, in particular, on all the studies I performed on the combinatorial background, that will be detailed in section 5.5.

This measurement has been performed on the full 2011 proton-proton collision data [106], corresponding to an integrated luminosity of 4.9 fb^{-1} and supersedes the previous one [107] made on data collected in the first part of 2011 and corresponding to an integrated luminosity of 2.4 fb^{-1}.

5.1 Analysis Methodology

The rare B decay analysis is based on events selected by the di-muon trigger (see section 2.3.1) and reconstructed in the ID (see section 2.2.3) and the MS (see section 2.2.5). The method adopted is based on a "cut & count" approach that uses a MultiVariate Analysis (MVA) classifier to discriminate between the $B_s^0 \to \mu^+\mu^-$ signal and the background events. The branching ratio is studied relative to a similar (only an extra track in the final state) and better known SM branching ratio with abundant observed statistics ($B^\pm \to J/\psi K^\pm$). This allows to partially cancel luminosity and efficiency uncertainties.

In order to avoid (experimenter sub-conscious) bias in the final result, a blind analysis is performed. This technique has been used in the past by a number of experiments including *BABAR* [108], BELLE [109], KTeV [110], E791 [111] and rare decay searches at BNL [112].

There are various approaches to blind analysis, that depend on the type of analysis itself. Typically, one has a Monte Carlo (MC) (or a control sample taken from data) to simulate the signal, and sidebands in data and/or a MC sample to characterise

the background to an analysis. Blind analysis means optimising all the cuts using such samples, testing the procedure to extract the signal yield, and evaluating the systematic errors before looking at the signal data. If a maximum likelihood fit is used, tests may even be performed on the signal data, but the fitted final results (and any plots which may reveal their approximate values) remain hidden until all checks have been made.

In this analysis all data in the invariant mass region [5066,5666] MeV are removed from the analysis until the procedure for the event selection, optimisation and limit extraction is completely defined and agreed upon.

Relative to the well established reference signal $B^{\pm} \to J/\psi(\to \mu^+\mu^-)K^{\pm}$, the $B_s^0 \to \mu^+\mu^-$ branching ratio can be written as:

$$\mathrm{BR}(B_s^0 \to \mu^+\mu^-) = \mathrm{BR}(B^{\pm} \to J/\psi K^{\pm}) \cdot \frac{(A_{J/\psi K^{\pm}} \cdot \varepsilon_{J/\psi K^{\pm}})}{(A_{\mu\mu} \cdot \varepsilon_{\mu\mu})} \cdot \frac{\mathcal{L}_{J/\psi K^{\pm}}}{\mathcal{L}_{\mu\mu}} \cdot \frac{f_u}{f_s} \cdot \frac{N_{\mu\mu}}{N_{J/\psi K^{\pm}}} \tag{5.1}$$

where, for each mode i ($i = \mu\mu, J/\psi K^{\pm}$), N_i is the number of observed events, ε_i and A_i are the absolute efficiencies and acceptances, and f_u/f_s is the relative production rate of $B^{\pm}$ and B_s^0 in the pp collisions. $\mathcal{L}_i$ are the integrated luminosities that in general can be different. The branching ratio for the reference channel was obtained from the PDG 2012 [121] as the product of $\mathrm{BR}(B^{\pm} \to J/\psi\, K^{\pm}) = (1.016 \pm 0.033)\ \times 10^{-3}$ and $\mathrm{BR}(J/\psi \to \mu^+\mu^-) = (5.93 \pm 0.06)\,\%$. The hadronisation fraction f_u/f_s is taken from the most updated experimental results available [123]: $f_s/f_d = 0.256 \pm 0.020$, considering $f_d/f_u = 1$. The dependence of the f_u/f_s on the decay kinematic was found to be negligible for this analysis [123].

The $N_{\mu\mu}$ yield reflects on a limit on the $\mathrm{BR}(B_s^0 \to \mu^+\mu^-)$ through a multiplicative coefficient referred to as Single Event Sensitivity (SES)

$$\mathrm{SES} = \mathrm{BR}(B^{\pm} \to J/\psi K^{\pm}) \cdot \frac{(A_{J/\psi K^{\pm}} \cdot \varepsilon_{J/\psi K^{\pm}})}{(A_{\mu\mu} \cdot \varepsilon_{\mu\mu})} \cdot \frac{\mathcal{L}_{J/\psi K^{\pm}}}{\mathcal{L}_{\mu\mu}} \cdot \frac{f_u}{f_s} \cdot \frac{1}{N_{J/\psi K^{\pm}}} \tag{5.2}$$

that can be seen as the branching ratio given by one single event

$$\mathrm{BR}(B_s^0 \to \mu^+\mu^-) = \mathrm{SES} \cdot N_{\mu\mu} \quad . \tag{5.3}$$

The SES can be effectively treated as an efficiency correction with uncertainty to the event yield $N_{\mu\mu}$. Since the triggers `EF_2mu4(T)_Bmumu` and `EF_2mu4(T)_Jpsimumu` were simultaneously active and not prescaled along the whole data taking period, we can synchronise the two yields and take $\mathcal{L}_{J/\Psi K^{\pm}}/\mathcal{L}_{\mu\mu} = 1$, obtaining:

$$\mathrm{SES} = \mathrm{BR}(B^{\pm} \to J/\psi K^{\pm}) \cdot \frac{(A_{J/\psi K^{\pm}} \cdot \varepsilon_{J/\psi K^{\pm}})}{(A_{\mu\mu} \cdot \varepsilon_{\mu\mu})} \cdot \frac{f_u}{f_s} \cdot \frac{1}{N_{J/\psi K^{\pm}}} \tag{5.4}$$

The $B^{\pm}$ yield (section 5.7) was obtained with an un-binned maximum likelihood fit after having applied a selection as similar as possible to the one used for the signal channel, in order to minimise the systematic uncertainties in the ratio. The terms $A \times \varepsilon$

(section 5.8) were evaluated using 2011 MC with systematic uncertainties estimated from the residual discrepancies between data and MC observed in the reference channel $B^{\pm} \to J/\psi\, K^{\pm}$.

The event selection is characterised by a baseline set of cuts followed by a final refinement based on a Boosted Decision Tree (BDT) that was trained on high-statistic MC background sample. The cut on the values of the BDT output was optimised on half of the sideband data. The number of signal events was estimated from the number of observed events in the signal region and the number of expected background events in the same region, where the background was interpolated from the still-unused half of sideband sample. The motivation of dividing the data mass sidebands in two parts is essentially related to the necessity of avoiding any kind of bias that could arise from using the same sample of events in both procedures.

5.2 Background Contributions

The expected background for the $B_s^0 \to \mu^+\mu^-$ analysis, given the 2011 sensitivity, can be subdivided into two categories: a continuum background that smoothly crosses the invariant mass search region, and all potential sources of resonance contributions.

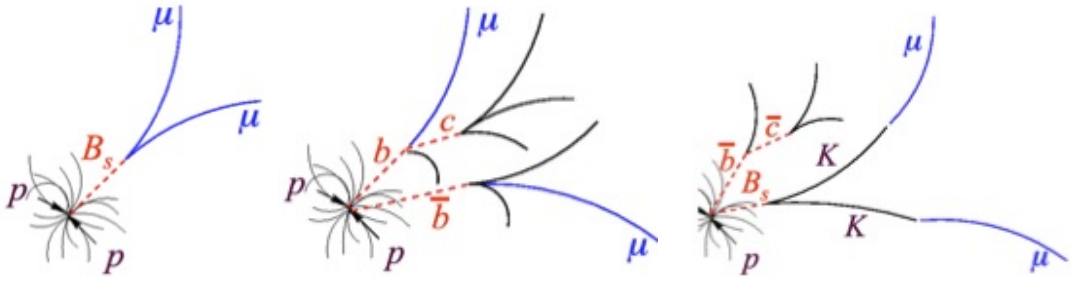


Fig. 5.1. *Event topologies contributing to this analysis. From left to right, $B_s^0 \to \mu^+\mu^-$ signal, continuum background, resonant background.*

Due to its smooth dependence on the di-muon invariant mass, the non-resonant background can be reliably modelled by the interpolation from the sidebands. This is the most prominent contribution to the analysis background and is made of processes $b\bar{b}(b\bar{b}b\bar{b}, b\bar{b}c\bar{c}) \to \mu^+\mu^- X$ with muons, produced mainly from semi-leptonic b and c decays, that can be randomly combined mimicking the kinematical signatures of the signal (see diagrams in figure 5.1). An inclusive $bb \to \mu^+\mu^- X$ MC sample ($\sim$ 200 million events) has been generated to model this contribution and showed good agreement with the sideband data once some basic kinematic re-weighting was applied. Since the event topology is different with respect to the signal one, we can define some discriminating variables to reject this continuum background.

The resonant background is made of B candidates containing one or two hadrons ($\pi^{\pm}$ or $\mathrm{K}^{\pm}$) that are erroneously misidentified as muons. This background is the most dangerous one since it peaks in the signal region (see left plot in figure 5.2) and has an event topology that is very similar to the signal one. Therefore, it constitutes a nearly

irreducible background in this analysis (ATLAS does not have a particle identification system). The mis-identification can be due to the punch-through of a hadron to the MS (the right side of figure 5.2 shows a simulation of this event in the ATLAS detector) or to decays in flight where the muon carries most of the hadron momentum. In the first case, the hadron fakes the muon signature ("fake muons")[1]. The probability of a hadron misidentified as muon in the 2011 analysis, studied using MC simulations, is found to be $2.1\,/\,4.1\,/\,3.3\,‰$ for $\pi^{\pm}\,/\,\mathrm{K}^{+}\,/\,\mathrm{K}^{-}$ respectively[2], with an uncertainty of 20 % [106, 113]. The expected event yield relative to the signal is obtained from the estimation of the integrated luminosity, acceptance and efficiency, and the latest measured branching ratios. The events obtained are included both in the optimisation procedure (section 5.6) and the extraction of the upper limit (section 5.10).

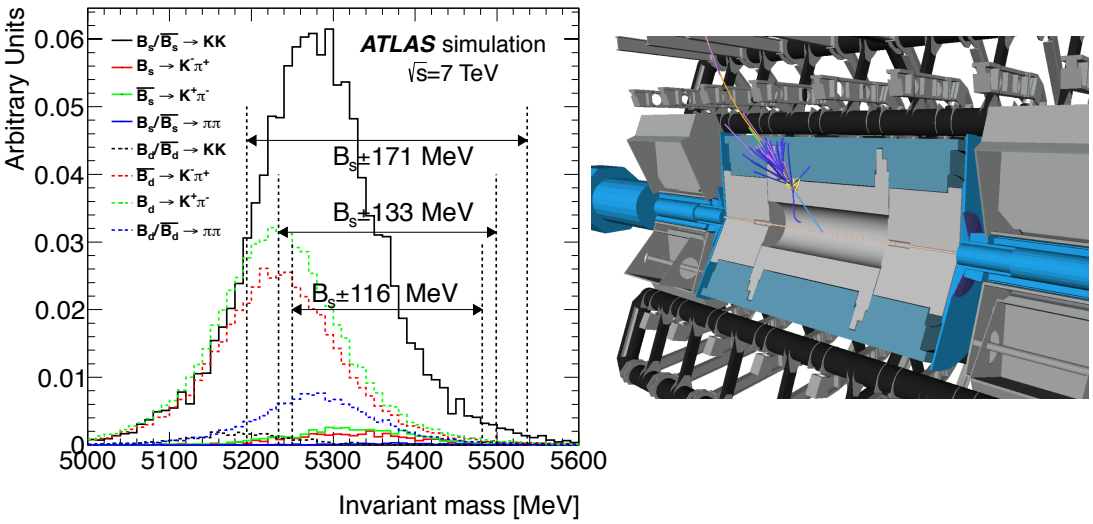


Fig. 5.2. *$B \to hh'$ $(h, h' = \pi^{\pm}, K^{\pm})$ invariant mass distribution with the separate resonant background contributions (left) and* GEANT *simulation of a muon punch-through generated by a high energetic charged π originating from the interaction point in the ATLAS detector [128] (right).*

5.3 Data and MC Sample

The 2011 analysis uses $\sqrt{s} = 7$ TeV pp collision data collected by ATLAS with stable LHC beam in 2011. Specific data-quality requirements have been applied in order to retain only the events in which both the tracking and muon detectors were fully operational. This leads to a total integrated luminosity of ~ 4.9 fb^{-1}.

[1] Even if the muons coming from the decay in flight are "true muons", from now on we will call both categories of events "fake muons".

[2] The misidentification probability is the ratio between the number of hadrons that have been identified as muons over the total number of generated hadrons.

The region of ± 300 MeV around the B_s^0 mass [5066-5666] MeV is "blinded" and, therefore, omitted from the analysis development. The sideband are defined in the intervals [4766,5066] MeV and [5666,5966] MeV respectively for the low and high sideband. Their width has been optimised in order to avoid mass ranges belonging to other possible di-muon resonances. The total mass window (no "blinding") for the $B^{\pm}$ reference channel is [4930,5630] MeV.

In 2011, after the September technical stop, the luminosity, and therefore the pile-up, was increased. Figure 5.3 reports the total recorded luminosity as a function of the mean number of interactions per bunch crossing (μ). In order to fulfil the strict trigger rate requirements, the L1 thresholds were increased. The 2011 data could be split into two sub-samples with roughly equal luminosity: in the first part the `EF_2mu4` chain was seeded at L1 from `L1_2MU0` in which no p_T cut on the muons was requested, while in the second part the chain `EF_2mu4T` was seeded from `L1_2MU4` in which a p_T cut at 4 GeV on the muons was introduced.

Tests on the discriminating variables were performed on the $B^{\pm} \to J/\psi\, K^{\pm}$ sample to study the impact of the two different triggers on the event yield. A negligible effect of $(1.1 \pm 1.5)\,\%$ difference between the event yields (normalised to the integrated luminosity) of the two chains, confirmed also in other ATLAS B-physics analyses, was observed. For that reason, the entire 2011 data has been considered as a whole consistent dataset.

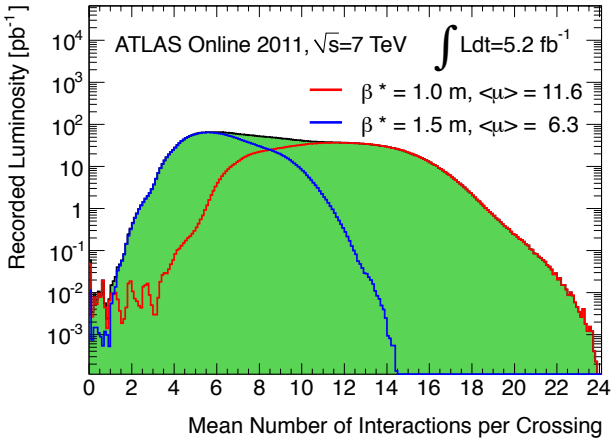


Fig. 5.3. *Luminosity-weighted distribution of the mean number of interactions per bunch crossing for the whole 2011 data taking period. After the September technical stop, the β^* was reduced from 1.5 m to 1.0 m. The mean number of interactions per crossing corresponds to the mean of the distribution of the number of interactions per crossing that is calculated from the instantaneous luminosity as $\mu = \mathcal{L} \cdot \sigma_{inel}/(n_{bunch} \cdot f_r)$ where $\mathcal{L}$ is the instantaneous luminosity, σ_{inel} is the inelastic cross section, which is taken to be 71.5 mb, n_{bunch} is the number of colliding bunches and f_r is the LHC revolution frequency [129].*

A series of MC samples, for the optimisation of the analysis and the extraction of the $A \times \varepsilon$ ratios, have been generated with PYTHIA 6.4 [114] using the ATLAS Monte

Carlo 2011 (MC11) tune [115]. The samples produced are summarised in table 5.1.

MC sample	Type	Events
$B_s^0 \to \mu^+\mu^-$	B_s^0 signal	$\sim$ 200 k
$B^0 \to \mu^+\mu^-$	B^0 signal	$\sim$ 200 k
$B^+ \to J/\psi\,(\mu^+\mu^-)\,K^+$	B^+ signal	$\sim$ 5 M
$B^+ \to J/\psi\,(\mu^+\mu^-)\,\pi^+$	B^+ resonant background	$\sim$ 1 M
$B_s^0 \to J/\psi\,(\mu^+\mu^-)\,\phi\,(K^+K^-)$	B^+ control sample	$\sim$ 730 k
$B \to hh'$	B_s^0 peaking background	$\sim$ 1 M
$b\bar{b} \to \mu^+\mu^- X$	B_s^0 combinatorial background	$\sim$ 200 M
$b\bar{b} \to J/\psi\,(\mu^+\mu^-)X$	B^+ combinatorial background	$\sim$ 15 M
$c\bar{c} \to \mu^+\mu^- X$	B_s^0 combinatorial background	$\sim$ 5 M
$pp \to \mu^+\mu^-$	B_s^0 Drell-Yan background	$\sim$ 500 k

Table 5.1. *MCs generated for 2011 analysis. The last two samples have been used for the studies reported in section 5.5.*

For all MCs (except $b\bar{b}$ samples) the ATLAS detector has been simulated using GEANT4 [116] in *full simulation* mode, in which all detectors are fully simulated. Since GEANT4 produces CPU-intensive tasks, given the very high statistic required, the two background $b\bar{b}$ samples were produced using Atlfast II [117], which employs the fast detector simulation for the calorimeters by means of parameterisations of the longitudinal and lateral energy profile, while the muon and tracking systems were fully simulated.

Every event produced has at least one decay of interest and it is selected to have all final state products generated within $|\eta| < 2.5$ and with the transverse momenta of both muons greater than 2.5 GeV. For the reference and control channels, the K^+ track can have p_T as low as 0.5 GeV. For the $b\bar{b} \to \mu^+\mu^- X$ continuum background sample, two muons with $p_T > 3.5$ GeV are required at generator level.

All MC signal samples have been tuned using an iterative re-weighting procedure: a Generator-Level (GL) re-weighting based on simulation, followed by a Data-Driven (DD) re-weighting. The first one is used to correct the biases in the relative $B_s/B^\pm$ fraction introduced by the selection applied at generator-level. For this purpose, dedicated MCs without selection on the final state and with a wider b-quark kinematic range have been generated. With these samples it is possible to create two binned (p_T^B, η^B) maps of the efficiencies for B_s and $B^\pm$ respectively. The inverse of such efficiencies are the weights that have to be applied to each event to correct for the GL filter biases. The DD re-weighting corrects for the residual (p_T^B, η^B) differences between data and MC observed after the GL re-weighting. The weights have been extracted by comparing $B^\pm \to J/\psi K^\pm$ data with MC events (the procedure is the

same used to correct the MC background $b\bar{b} \to \mu^+\mu^- X$ and is detailed in section 5.5). In order to not correlate the re-weighting procedure with the reference channel yield extraction, only candidates with odd event numbers in the ATLAS dataset are used to determine the weights, while the remaining sample is used for the yield measurement. The weights obtained have been cross-checked on $B_s^0 \to J/\psi\,\phi$ control channel.

In addition, since PYTHIA does not reproduce angular distributions in the $B^\pm \to J/\psi\,K^\pm$ and $B^\pm \to J/\psi\,\pi^\pm$, these MC samples have been corrected a posteriori applying per-event weights that allow to reproduce the correct J/ψ longitudinal polarisation.

5.4 Candidate Preselection

A baseline selection, based on the guidelines of the corresponding ATLAS performance subgroups, is applied to all MCs and data samples in order to provide a first skim of the events. All events are selected using the di-muon trigger `EF_2mu4(T)_Bmumu` or `EF_2mu4(T)_Jpsimumu` for the signal and reference channel respectively (see sections 2.3.1, 2.3.2 for more details), requiring two muons with p_T above 4 GeV.

Following the recommendations of the ID and muon tracking subgroups, good-quality reconstructed tracks are selected requiring at least 1, 6 and 9 hits[3] in the Pixel, SCT and TRT detectors respectively, with a number of holes[4] in the Pixel+SCT lower than 3. In addition, in order to increase the tracking precision, 1 hits in the pixel innermost layer (b-layer) is required.

All B candidates are demanded to satisfy a request on the transverse momentum $p_T^B > 8$ GeV and on the pseudo-rapidity $|\eta| < 2.5$. In addition, the $B \to \mu^+\mu^-$ candidates (signal and $b\bar{b}$ background) are selected requiring two combined muons with $p_T > 4$ GeV and $|\eta| < 2.5$, with the common vertex fulfilling the condition $\chi^2/\mathrm{NDF} < 2$. The same selection on the muons are applied to $B^\pm \to J/\psi\,K^\pm$ and $B_s^0 \to J/\psi\,\phi$, but with the additional requirement of having, respectively, one or two further tracks with $p_T > 2.5$ GeV and $|\eta| < 2.5$ (in both cases, the three or four tracks common vertex should satisfy $\chi^2/\mathrm{NDF} < 6$)[5].

Table 5.2 shows the efficiency of the baseline selection with respect to the fiducial phase-space volume ($p_T^B > 8.0$ GeV and $|\eta^B| < 2.5$) for the $B_s^0 \to \mu^+\mu^-$ and $B^\pm \to$

[3] The hit efficiency in the silicon detector is close to 100 % and only a few modules were inactive during data taking. The transverse momentum is the main parameter that determines how far the particle can reach out into the tracking volume, or, in other words, how long the reconstructed track is expected to be (it also depends on other kinematic parameters of the particle: its production vertex and the direction of the initial particle momentum, since the geometrical detector setup allows different numbers of crossed detection layers and thus leads to a different number of average hits). A shorter track length results in a worse momentum resolution that leads to the effect of low p_T particles to be measured as high p_T tracks. To suppress this migration effect, for a $p_T > 300$ MeV a track is required to have, at least, 6 hits in the SCT detector. TRT extensions require a minimum of 9 hits in the TRT, at least 50 % of which are precision hits.

[4] A hole is defined as an ID layer in which an offline track has a hit, but the matched L2 track does not, i.e. a missing measurement when it is expected.

[5] This selection corresponds to an acceptance of 99.5 %.

$J/\psi K^{\pm}$ channels evaluated in data and MC 2011.

Channel	2011 MC	2011 data
$B_s^0 \to \mu^+\mu^-$	20.5 %	18.2 %
$B^{\pm} \to J/\psi K^{\pm}$	4.9 %	8.2 %

Table 5.2. *Efficiency of the baseline selection with respect to the fiducial phase-space volume ($p_T^B > 8.0$ GeV and $|\eta^B| < 2.5$) for $B_s^0 \to \mu^+\mu^-$ and $B^{\pm} \to J/\psi K^{\pm}$.*

5.5 Studies on the Combinatorial Background

In the past analysis, performed only on the first half of the 2011 data taking period, the sideband data were used for the training of the multivariate discriminating variable (BDT), the optimisation of the selection and the interpolation of the background to the signal region. For this procedure, the sample was divided into even- and odd-numbered events, half statistic was devoted to the background interpolation and the other half for the rest of the analysis. Even thought it has been demonstrated that the bias introduced by this procedure is rather small, ideally, one would have the possibility to perform each step on different samples, in order to avoid any possible mass correlation.

The idea for the 2011 analysis on 4.9 fb^{-1} of data was to use the $b\bar{b} \to \mu^+\mu^- X$ MC sample to train and test the MVA leaving the whole sidebands available for the optimisation of the selection cuts and the interpolation of the background to the signal region. Before starting to use this MC, it is necessary to test the reliability of this sample in correctly modelling, in terms of the discriminating variables, the sideband data that describe the real background expected in our analysis.

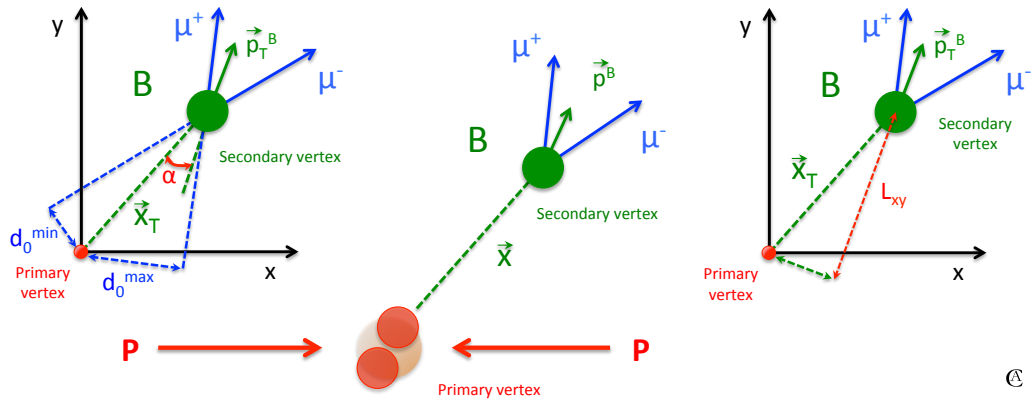


Fig. 5.4. *Illustration of some of the discriminating variables used in the analysis and reported in table 5.3. Left: pointing angle α_{2D}, d_0^{min}, d_0^{max}. Right: L_{xy}.*

Variable	Description
$\alpha_{\,2D}$ (pointing angle)	Angle in the transverse plane between the vector from the primary to the secondary vertex $\Delta\vec{x}$ and the B_s^0 momentum $\vec{p}^{\,B}$
ΔR	Angular distance $\sqrt{\Delta\phi^2 + \Delta\eta^2}$ between $\Delta\vec{x}$ and $\vec{p}^{\,B}$
L_{xy}	Scalar product in the transverse plane of $\Delta\vec{x} \cdot \vec{p}^{\,B}/\lvert\vec{p}_T^{\,B}\rvert$
Proper time significance	Proper decay length $ct = L_{xy} \times m_B/p_T^B$ divided by its uncertainty (the values of m_B and p_B are obtained from the refitted track momenta given by the vertex-fit for the fully reconstructed candidate)
$\log(\chi^2_{xy})$, $\log(\chi^2_z)$	Vertex separation significance between PV and SV in the x-y plane and along the z-axis respectively
$I_{0.7}$ (isolation)	Ratio of $\lvert\vec{p}_T^{\,B}\rvert$ to the sum of $\lvert\vec{p}_T^{\,B}\rvert$ and the transverse momenta of all tracks with $p_T > 0.5$ GeV within a cone $\Delta R < 0.7$ from the B direction, excluding the B decay products
$\lvert d_0^{\min}\rvert$, $\lvert d_0^{\max}\rvert$	Absolute values of the minimum and maximum impact parameter in the transverse plane of the B decay products relative to the primary vertex
DCA, ZCA	Values of the minimum distance of closest approach of tracks in the event with respect to the B vertex, in the x-y plane and along the z-axis respectively
p_L^{min}	Minimum momentum of the two muon candidates along the B direction
p_T^B	B transverse momentum

Table 5.3. *List of the 13 discriminating variables used to reject the continuum background in the 2011 analysis. These variables are based on properties of the decay products, of the reconstructed primary ($\vec{x}_{PV}$) and secondary ($\vec{x}_{SV}$) vertices (separated by $\Delta\vec{x} = \vec{x}_{SV} - \vec{x}_{PV}$), the B meson momentum $\vec{p}_T^{\,B}$ and the properties of additional tracks from underlying events. To minimise the pile-up dependence, the isolation is evaluated only considering tracks originating from the PV associated to the B decay.*

5.5.1 Discriminating Variables

The strategy for the $B_s^0 \to \mu^+\mu^-$ selection is based on its topology which is quite simple. Given its long lifetime ($\tau = 1.47 \pm 0.03$ ps), after the production at the Primary Vertex (PV) the B_s^0 meson travels an appreciable distance ($c\tau = 441$ μm) before decaying. The muons stemming from the decay vertex (Secondary Vertex, SV) are isolated and the sum of their transverse momenta lies at a small angle with

respect to the B_s^0 line on flight. To minimise the pile-up dependence, the isolation is evaluated only considering tracks originating from the PV associated to the B decay (this requirement makes the selection mostly independent of pile-up as shown in figure 5.5). All these features allow to discriminate the signal sample from the background. For example, the Drell-Yan processes ($pp \to \mu\mu$) can be easily rejected by means of the variables related to the SV, due to the fact that in this case the true muons come directly from the interaction point.

The discriminating variables have been chosen among a large number of topological event-shape variables on which detailed studies, checking the mass correlation and the discriminating power, were done in preparation of the analysis performed on the first part of 2011 data taking period. The list was finally restricted to 14 variables.

For 2011 measurement the same selection was adopted, except for one variable (the maximum momentum of the two muon candidates along the B direction) that, not significantly improving the background rejection performance, was not considered in the analysis. Therefore, no new variables were investigated. The list and a brief description of the 13 variables is reported in table 5.3.

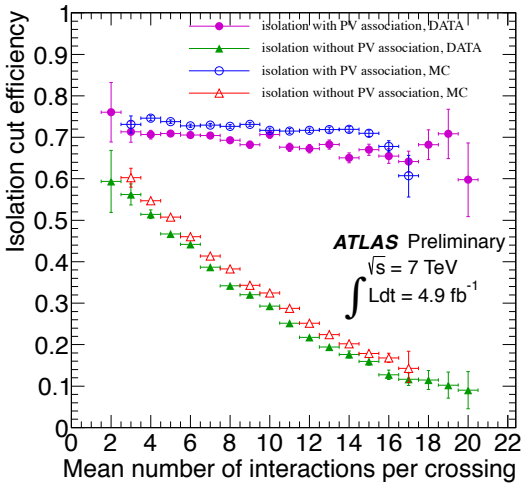


Fig. 5.5. *Efficiency of the isolation cut $I_{0.7} > 0.83$ as a function of the pile-up for $B^{\pm} \to J/\psi\, K^{\pm}$ candidates from data (filled symbols) and MC (empty symbols) [107]. Missing MC points at the edges of the distributions are due to the lack of statistics in the MC simulation.*

5.5.2 Test on the 2010 data and MC

Preliminary tests on the discriminating variables were performed comparing the 2010 sideband data (~ 45 pb^{-1}) with the 2010 MC $b\bar{b} \to \mu^+\mu^- X$ in order to check the reliability of the simulation. The MC background samples used are:

- $b\bar{b} \to \mu^+\mu^- X$ with *no pile-up (full simulation)*: $\sim$ 775,878 generated events corresponding to an integrated luminosity of 1.6 pb^{-1}.

- $b\bar{b} \rightarrow \mu^+\mu^- X$ with *pile-up (full simulation)*: $\sim$ 5,809,031 generated events corresponding to an integrated luminosity of around 12 pb^{-1}.

- $b\bar{b} \rightarrow \mu^+\mu^- X$ with *pile-up (fast simulation)*: $\sim$ 200 K generated events processed with Atlfast II (0.4 pb^{-1} of equivalent luminosity).

The events considered are the ones that pass the baseline selection reported in section 5.4. Implementing the final selection, based on the discriminating variables, leaves essentially no events in the MC and few events in data. As we can see in figure 5.6, that reports only the most powerful discriminating variables, due to the very low statistic, we can not draw a conclusion about the reliability of the MC to model the sideband data using the 2010 samples.

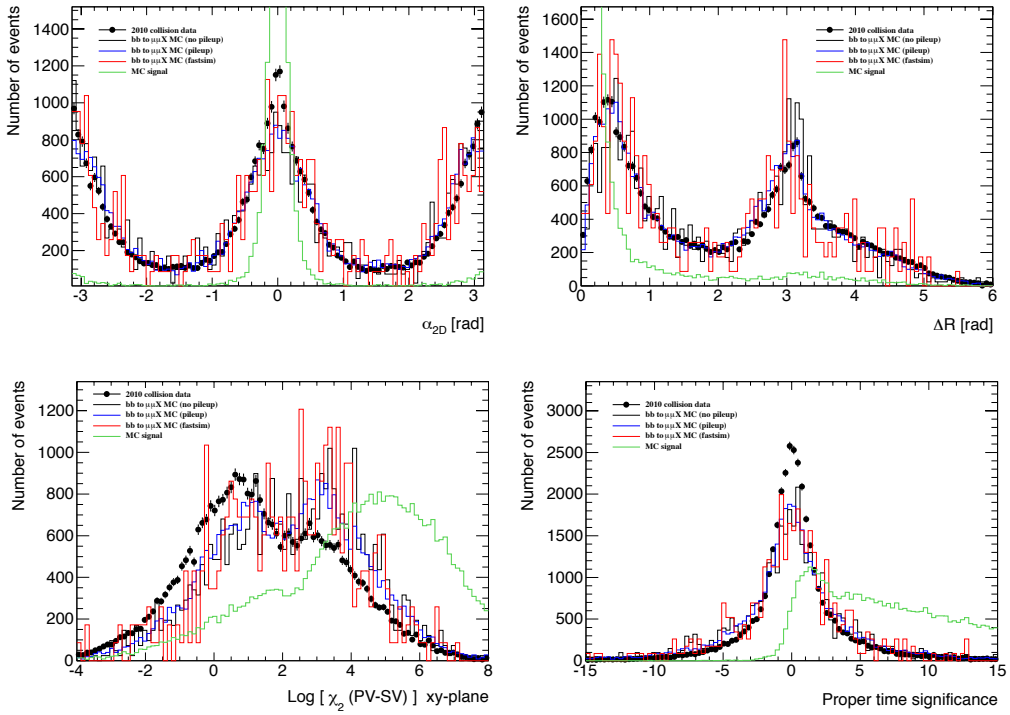


Fig. 5.6. *Data-MC comparison for a selection of the most powerful discriminating variables using the 2010 data and the MC generated events (the event selection applied is the one reported in section 5.4). The black dots are the 2010 sideband data, the black, blue and red histograms represent respectively the simulated event with no pile-up, with pile-up and the MC processed with Atlfast II. The green line is the MC signal, reported for shape comparison.*

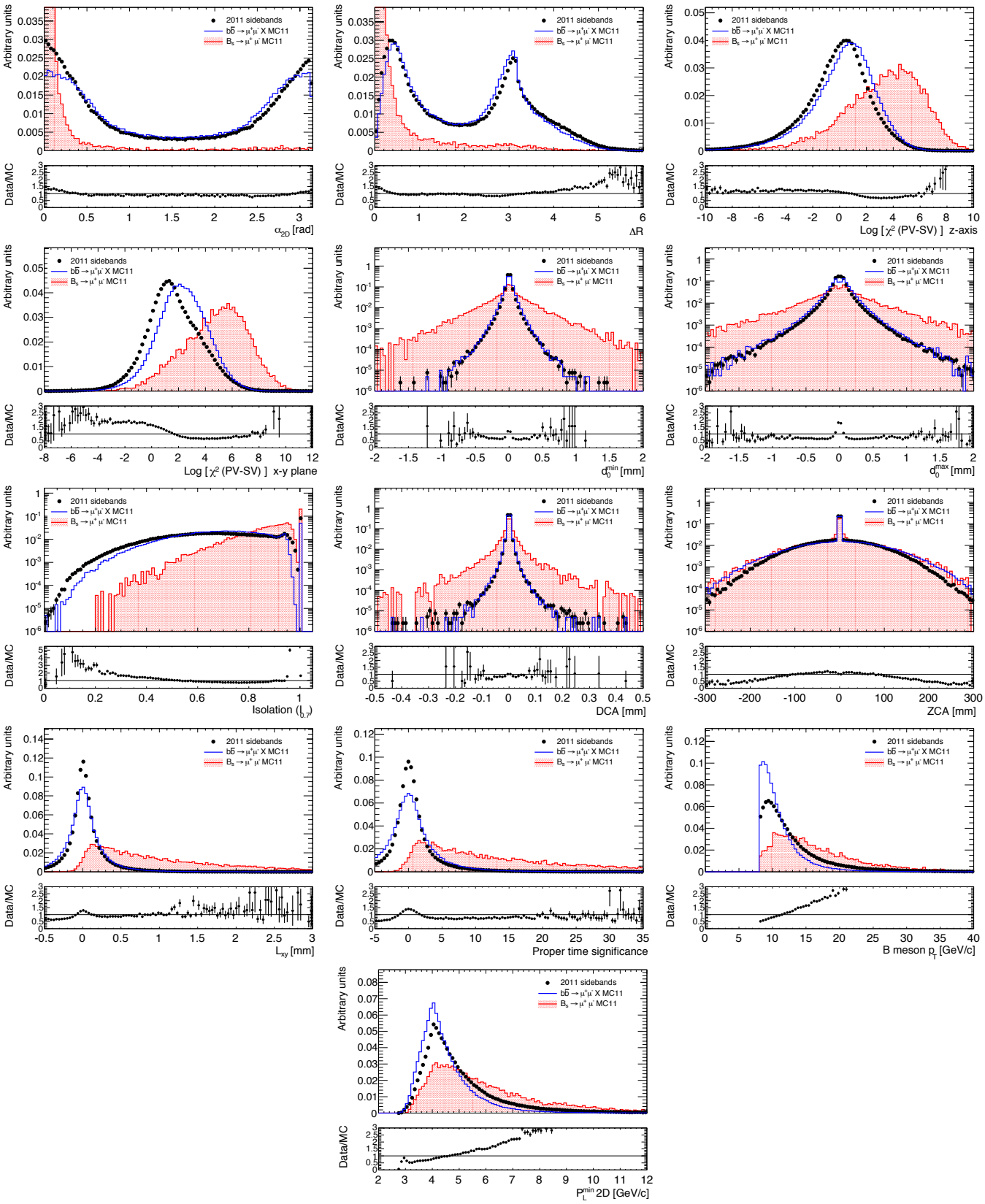


Fig. 5.7. *Distributions of the 13 discriminating variables used in the analysis. The black dots represent the 2011 sideband data, the blue solid histogram is the $b\bar{b} \to \mu^+\mu^- X$ MC processed with Atlfast II. For shape comparison, the signal MC is shown (red-filled solid histogram). For each plot it is also reported the ratio between data and MC distributions.*

5.5.3 Test on the 2011 data and MC

More detailed checks can be performed on the whole 2011 data and the 2011 MC sample $b\bar{b} \to \mu^+\mu^- X$ with 200 M events that correspond to an integrated luminosity of ~ 0.7 fb^{-1} (the cross-section of the $b\bar{b}$ processes, where the two muons are generated with a p_T higher than 3.5 GeV, is taken to be ~ 270 nb). Figure 5.7 shows the data-MC comparison for all the 13 discriminating variables. The discrepancies between the data and the MC prove how this sample, at this stage, could not be reliably used to simulate the non-resonant background of the sidebands.

Several studies have been performed trying to understand the possible sources of these discrepancies. Inaccuracies can be expected in the fast simulation techniques that can introduce biases in the detector simulation. Indeed, Atlfast II uses the parameterisation of the calorimeters, but the ID and the MS are fully simulated as for GEANT4. This can be tested comparing the distributions of some fundamental variables like the positions of the PV and B vertices, and the kinematic variables of the muons, obtained with the full and fast simulation. For this purpose, we can analyse the datasets used for the preliminary tests (section 5.5.2). The small test sample (~ 200 K events) $b\bar{b} \to \mu^+\mu^- X$ MC10 processed with Atlfast II is compared with $b\bar{b} \to \mu^+\mu^- X$ MC10 processed with the full GEANT4 simulation (~ 6 M events). Figure 5.8 shows the distributions of PV and SV position in the transverse plane, as well as η and p_T of the muon tracks for full simulation, fast simulation and 2010 sideband data. As expected, the fast simulation is in a very good agreement with the full simulation, while both differ from the data distributions.

A second source of possible discrepancies can be explained by the presence in data of background contributions that have not been simulated in the MC and that, having a different kinematic with respect to the continuum, have a different behaviour in terms of discriminating variables. At low transverse momentum, muons can come from prompt Drell-Yan (DY) processes, beauty and charm hadron decays, and light meson decays. We already know that the background from $b\bar{b}$ events is the dominant one and that the resonant background, where the hadrons are misidentified as muons, only contributes to the signal region. Therefore, we need to focus only on the di-muons from DY and $c\bar{c}$ decays.

A dedicated sample of low-mass DY MC (generated for charmonium analyses) and $c\bar{c} \to \mu^+\mu^- X$ have been used for this purpose (see table 5.1). Applying the baseline selection (section 5.4), only a very small fraction (< 0.3 %) of DY pairs survives the di-muon trigger cuts in the interested mass range, which makes this contribution essentially negligible. Charm decays, despite having an estimated total rate twice as high as from beauty, contribute at a much lower level (< 2 %) than b, due to the p_T distribution of the muons from c quark that falls more steeply.

We conclude that the inconsistencies found in the $b\bar{b} \to \mu^+\mu^- X$ MC are related to discrepancies in the physics modelling at generator level (e.g. B meson kinematics) and residual inaccuracies in the detector description. The standard procedure adopted in these cases is to tune the kinematics of the MC sample from data, introducing per-event weights.

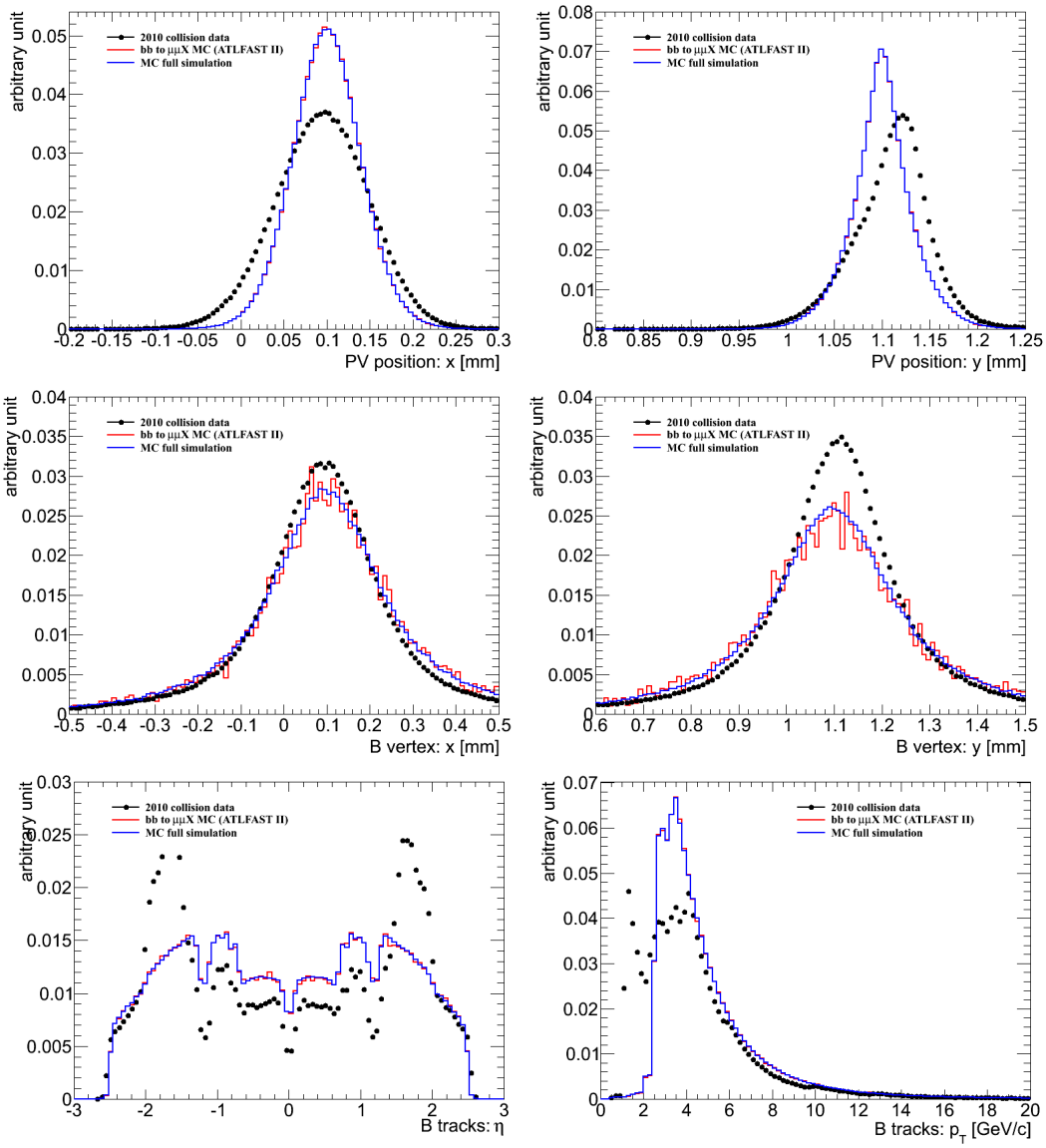


Fig. 5.8. *Distributions of x and y position of PV and B vertices, as well as η and p_T of the muon tracks for 2010 sideband data (black points), MC Atlfast II (red line) and MC* GEANT4 *(blue line).*

5.5.4 Data Driven Weight for the Background Sample

Several tests have been performed on different variables (position of PV and B_s^0 decay vertices, as well as ϕ, η and p_T of the muons), re-weighting the MC sample in terms of them and checking the effect on the discriminating variables distributions. The search has been finally restricted to the transverse momentum p_T^B and the pseudo-rapidity η^B of the di-muons candidates, whose modelling inaccuracies significantly contribute to the data-MC discrepancies of the discriminating variables.

In addition, among all fundamental variables analysed, and that can be used in the re-weighting procedure, the number of primary vertices is one of the most important. A priori, the number of interactions per bunch crossing, and consequently the number of primary vertices in data, could differ from the MC. Figure 5.9 shows the distributions of the number of reconstructed primary vertices (left) and the number of interactions per bunch crossing μ (right) for the first and the second part of the 2011 data taking, showing the differences in the pile-up between the two periods. Since the MC is in a very good agreement with the sideband data, the PV has not been included in the re-weighting procedure.

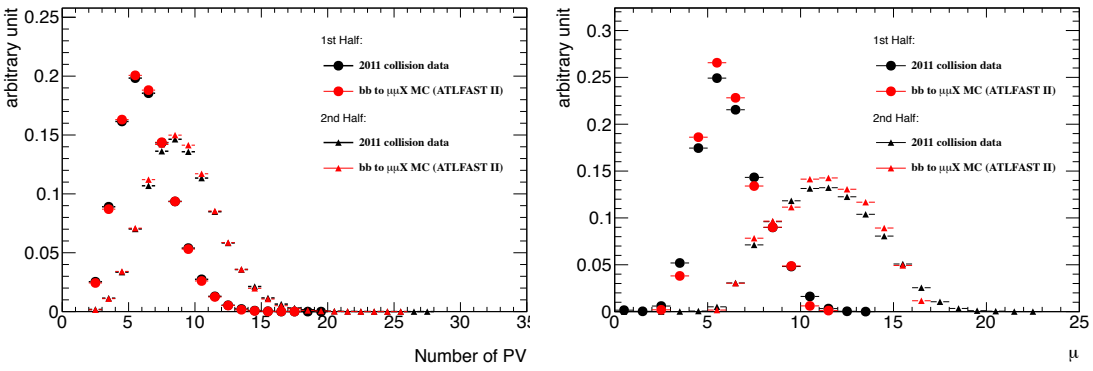


Fig. 5.9. *Distributions of the number of primary vertices (left) and μ (right) for the first and the second half of the 2011 data taking period. The black dots and triangles are the sideband data whereas the red ones are the $b\bar{b} \to \mu^+\mu^- X$ MC.*

Two sets of DD weights have been determined by the ratio of the normalised 1D p_T^B and η^B spectra in the sideband data and the MC. The total weights can be expressed as

$$W(p_T^B, \eta^B) = w(p_T^B) \cdot w(\eta^B) \ , \qquad \text{where} \quad w(y) = \frac{n_i^{\text{data}}(y)}{n_i^{\text{MC}}(y)} \tag{5.5}$$

and n_i is the normalised number of entries in either the sideband data and the MC histograms for p_T^B and η^B.

In order to reduce the correlation between the weights and the sideband data and, therefore, avoid possible biases, the data distributions of the p_T^B and η^B have been parameterised [6] using respectively a 4^{th} degree polynomial plus a *Cruijff function*[7], the latter defined as

$$\mathcal{C} = \begin{cases} A \cdot exp\left(-\frac{(x-m)^2}{2\sigma_L^2+\alpha_L(x-m)^2}\right) & x < m \\ A \cdot exp\left(-\frac{(x-m)^2}{2\sigma_R^2+\alpha_R(x-m)^2}\right) & x \geqslant m \end{cases} \ , \tag{5.6}$$

[6] The standard procedure is to divide, bin-by-bin, the bin content of the data distribution by the bin content of the MC distribution.

[7] The Cruijff function is a centred Gaussian with different left-right resolutions and non-Gaussian tails. The name is a tribute to Hendrik Johannes Cruijff, former Dutch footballer.

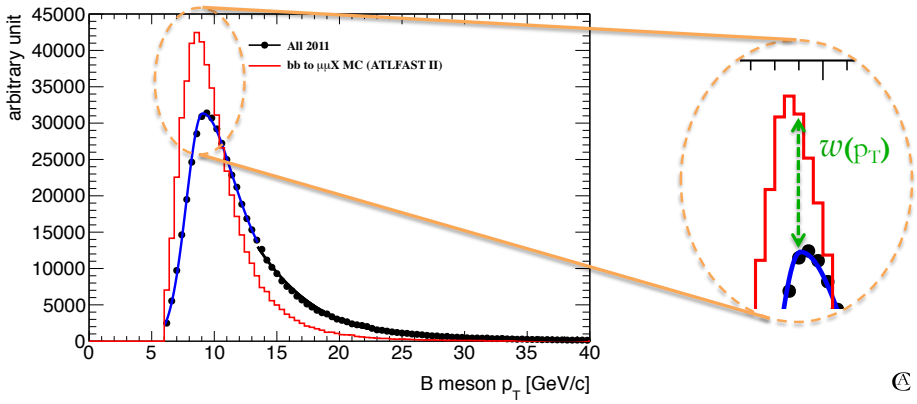


Fig. 5.10. *This figure shows how the weights (p_T^B in this case) are evaluated: for a certain p_T bin of the two distributions, we need to rescale the red solid line to the blue and black curves, computing the scaling factor $w(p_T)$. This can be done taking the value of the parameterised function corresponding to the p_T^B of the centre of the considered MC bin, and dividing the obtained value by the number of entries contained in the MC bin (the MC is normalised to data).*

and two 7^{th} order polynomials

$$P_n(x) = c_0 + \sum_{i=1}^{7} c_i \cdot x^i \quad , \tag{5.7}$$

one for $\eta \geqslant 0$ and the other for the negative values[8]. The results of the fits are reported in the figure 5.11 and summarised in the tables 5.4 and 5.5. The weights, obtained as explained in the figure 5.10, are used as per-event weight on the MC and the whole procedure is iterated until they stabilise (only if the weights converge to the unity the method can be considered reliable).

The procedure has been iterated twice and the total weights (reported in the figure 5.13) are the product of the ones obtained in the first and the second iteration. Finally, the convergence of this method has been checked, and after the second iteration (see figure 5.14) the weights are almost equal to the unity.

The DD weights for the $b\bar{b} \to \mu^+\mu^- X$ MC have been evaluated analysing the whole 2011 data. Possible effects related to the two different triggers present in the first and second part of the data taking have been studied repeating the whole procedure on each of the two sub-samples. The weights obtained, reported in figure 5.12, do not show any significant difference between the two periods, confirmed by the detailed tests performed on the trigger (see section 5.3).

Once the p_T^B and η^B weights are applied, small discrepancies are still present (see figures 5.15 - 5.21). Since the purpose of this MC is the training of the BDT, in order to avoid a suboptimal classifier a good data-MC agreement is required in the parts of the discriminating variable distributions that have a signal-like behaviour and where

[8] Since the distribution is asymmetric with respect to $\eta = 0$.

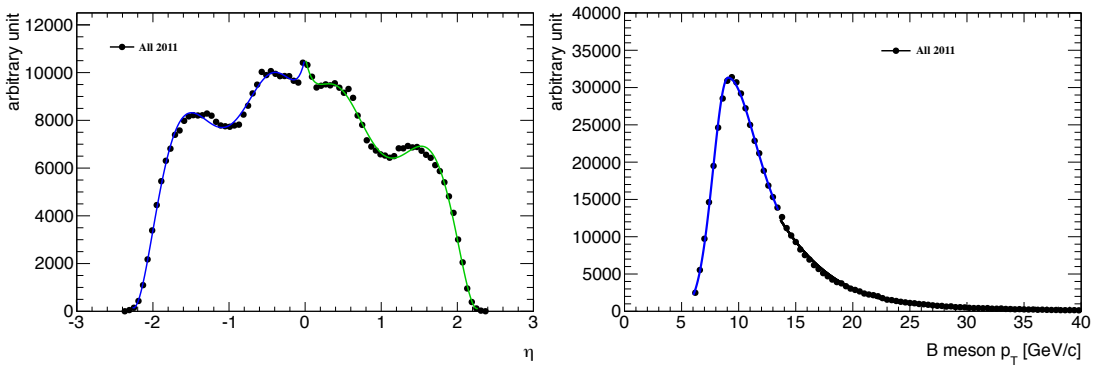


Fig. 5.11. *B_s^0 meson η (left) and p_T (right) parameterisation obtained with respectively two 7^{th} order polynomials and a Cruijff function plus a 4^{th} degree polynomial.*

pol 7^{th}: $\eta < 0$	
χ^2/NDF	148/30
$c0$	10634 ± 135
$c1$	15666 ± 2030
$c2$	88465 ± 9768
$c3$	199308 ± 20910
$c4$	201366 ± 23025
$c5$	97081 ± 11566
$c6$	20777 ± 2491
$c7$	1408 ± 152

pol 7^{th}: $\eta > 0$	
χ^2/NDF	215/30
$c0$	10663 ± 134
$c1$	- 15936 ± 1389
$c2$	79429 ± 9450
$c3$	- 174008 ± 20108
$c4$	169735 ± 18044
$c5$	- 77519 ± 8946
$c6$	15042 ± 1368
$c7$	- 775 ± 99

Table 5.4. *results of the η^B fit.*

Cruijff	
χ^2/NDF	186/13
m	9.06 ± 0.02
σ_L	1.30 ± 0.01
σ_R	2.80 ± 0.05
α_L	0.0049 ± 0.0002
α_R	0.38 ± 0.01
A	31328 ± 91

pol 4^{th}	
χ^2/NDF	906/64
$c0$	94989 ± 256
$c1$	- 1504 ± 93
$c2$	529 ±14
$c3$	- 10.8 ±0.6
$c4$	0.083 ± 0.001

Table 5.5. *results of the p_T^B fit.*

therefore the MVA cut is optimised. The previous statement can be proved selecting the B_s^0 candidates with a decay length (L_{xy}) higher than 200 μm (that retains only the part of the sample which is more similar to the signal), and verifying that the

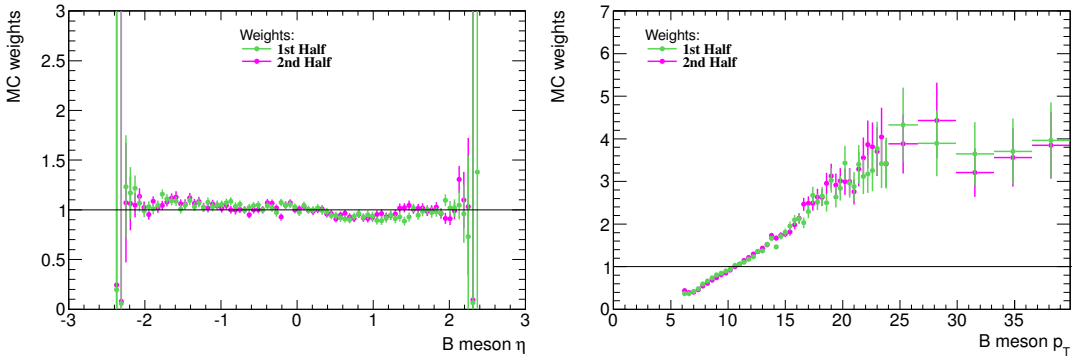


Fig. 5.12. *Comparison of the B_s^0 meson η (left) and p_T (right) weights evaluated for the 1^{st} and the 2^{nd} half of the 2011 data taking period.*

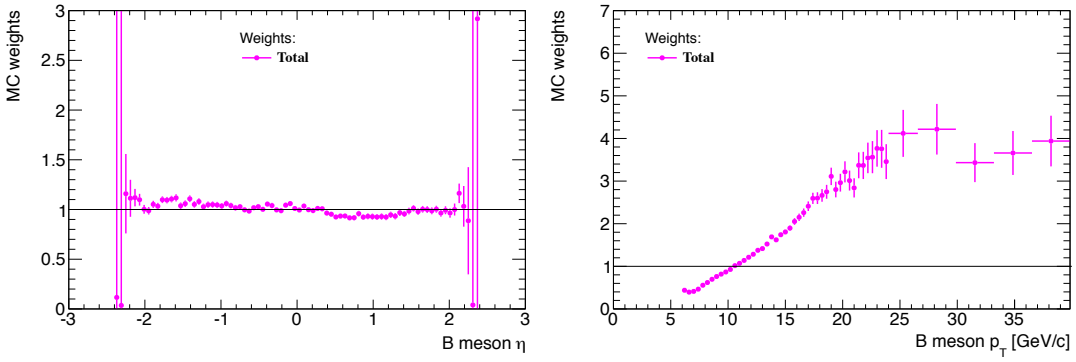


Fig. 5.13. *B_s^0 meson η (left) and p_T (right) weights obtained by the product of the first and second iteration.*

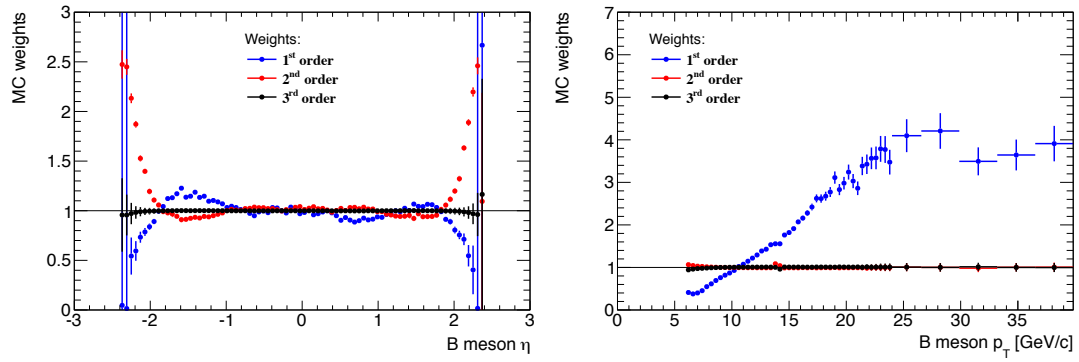


Fig. 5.14. *B_s^0 meson η (left) and p_T (right) weights obtained with the 1^{st}, 2^{nd} and 3^{rd} iteration. The method clearly converges to unity.*

agreement data-MC of the distributions improves. The final results are reported in the figures 5.15 - 5.21 and prove how the data-MC discrepancies are reduced in the

signal-like region.

The studies reported in this section show how the high statistic $b\bar{b} \to \mu^+\mu^- X$ MC sample can be used to model the background present in the sideband data after having applied the per-event DD weights. Unlike the past analysis, that used the odd-numbered sideband data events for the multivariate analysis, for the analysis on the full 2011 data it was possible to reliably base the training and testing procedure of the MVA on the MC background.

The 13 discriminating variables were all included in a MVA classifier, using the TMVA package [118]. Several classifiers were compared and the best ones were found to be those based on the BDT algorithm. Several configurations of the BDT have been tested and a specific optimisation has been studied to choose the best performing one (see section 5.6).

5.6 Signal Selection Optimisation

The optimisation of the selection, necessary to efficiently extract the signal from the background events, can be essentially summarised in three steps. First of all we need to choose the estimator to be used for the ranking of the selection strategies, then train a number of BDT configurations on MC events, and finally optimise the BDT and the mass window selection on half sideband events. The optimal value corresponds to the maximum value of the estimator in the (BDT , Δm) space where Δm is the width of the search window centred around the B_s^0 mass fixed to 5366.33 MeV for this scope.

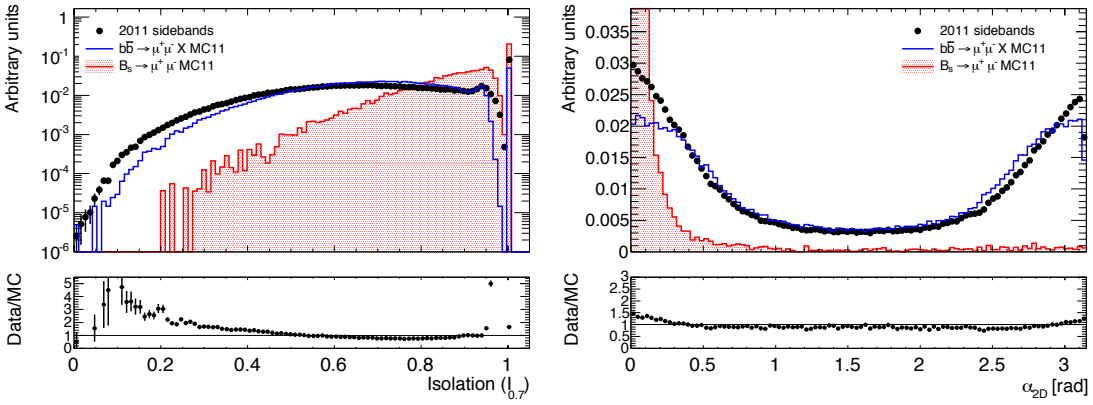
The performance of the various strategies are ranked by means of the Punzi [119] estimator:

$$\mathcal{P} = \frac{\varepsilon_{\,\text{signal}}}{\frac{a}{2} + \sqrt{N_{\,\text{bkg}}}} \tag{5.8}$$

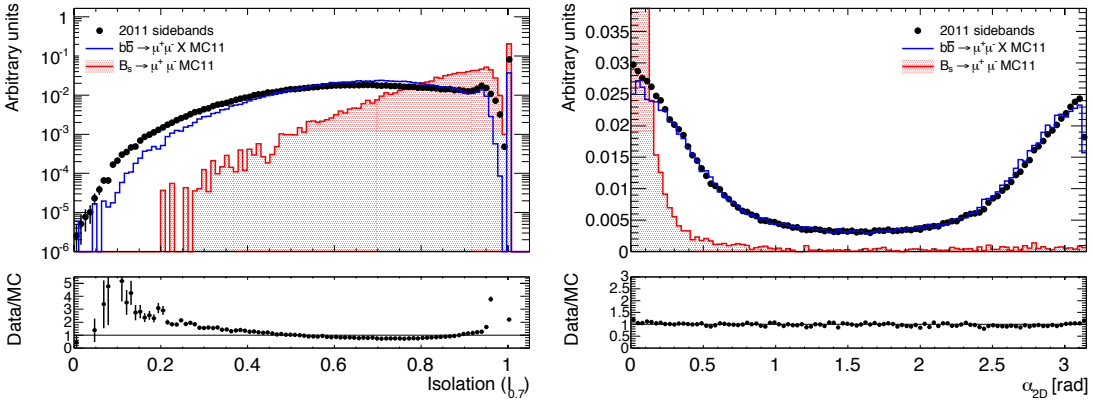
where $\varepsilon_{\,\text{signal}}$ and $N_{\,\text{bkg}}$ are the signal efficiency and the background yield respectively for a given configuration, while a is function of the aimed confidence level in the analysis ($a = 2$ for 95 % of CL). The choice of this estimator [9] is related to our lack of knowledge of the $B_s^0 \to \mu^+\mu^-$ branching ratio and it is specifically designed to maximise the performance of a cut & count frequentist limit. Effectively, we are optimising for two sigma background exclusion and two sigma signal discovery. The 2-dimensional optimisation on the BDT output requirement and the signal region width is performed on the signal MC sample and the odd-numbered data sideband events.

Figure 5.22 reports the distributions of the chosen BDT output variable for signal MC events and sideband data. The odd-numbered event optimisation gives a maximum $\mathcal{P}$ value of 0.0145 that corresponds to a selection on the BDT output > 0.118 and on the mass windows $|\Delta m| < 121$ MeV. The optimised BDT and mass windows selections have a signal efficiency equal to 25 % and 84 % respectively. The efficiency of the combination of the two selections is therefore 21 %.

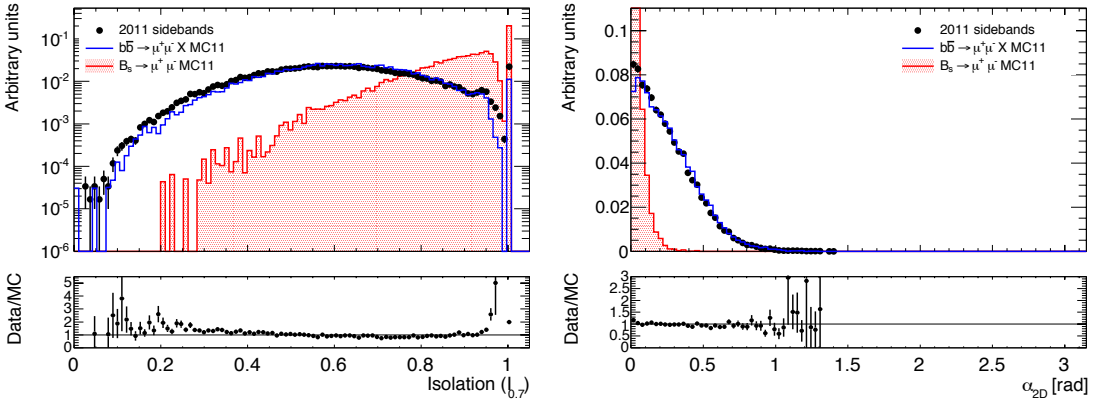
[9] The most common estimators are the well known $\frac{S}{\sqrt{S+B}}$ and $\frac{S}{\sqrt{B}}$.



(a) *No re-weighting is applied.*

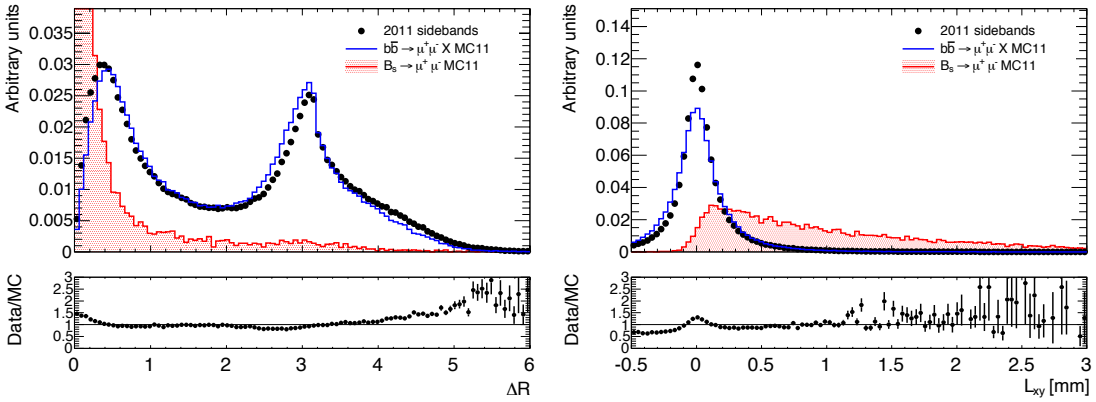


(b) *Re-weighting to the sidebands p_T and η distributions is applied.*

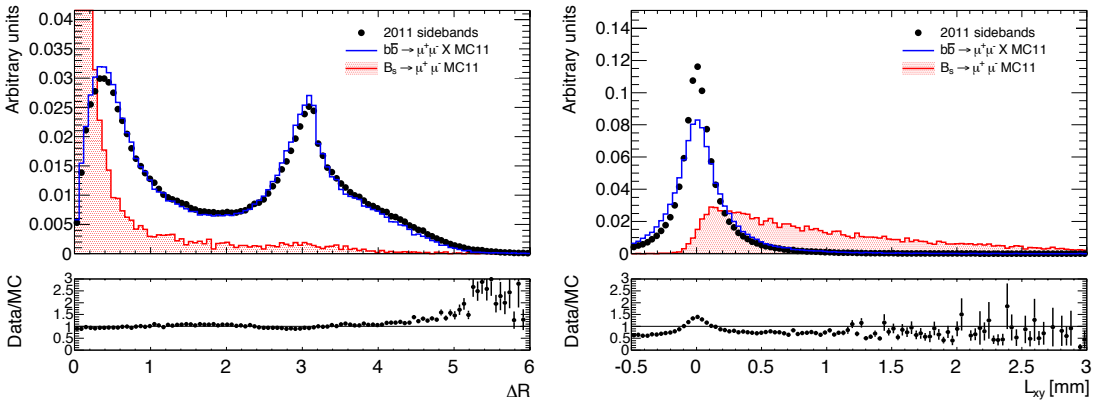


(c) *Re-weighting to the sidebands and a $L_{xy} > 0.2$ mm cut are applied.*

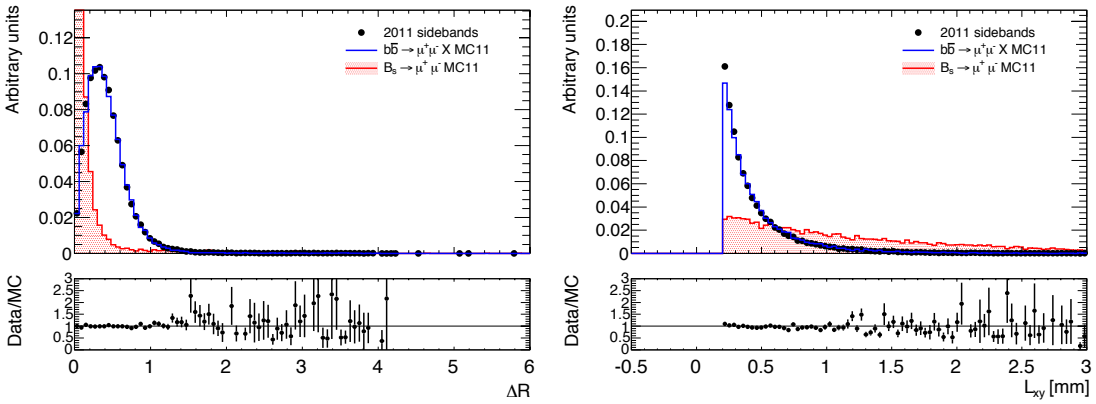
Fig. 5.15. *Distributions of Isolation (left), and pointing angle in 2D (right). The sideband data (black dots) are compared with the $b\bar{b}$ background MC (blue solid histogram). The signal is shown for shape comparison (red-filled solid histogram). For each plot it is also reported the ratio between data and MC distributions.*



(a) *No re-weighting is applied.*

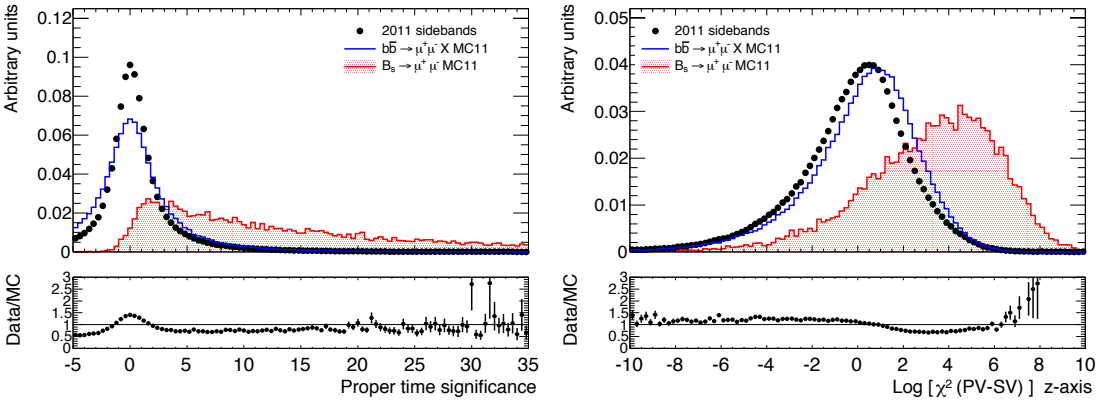


(b) *Re-weighting to the sidebands p_T and η distributions is applied.*

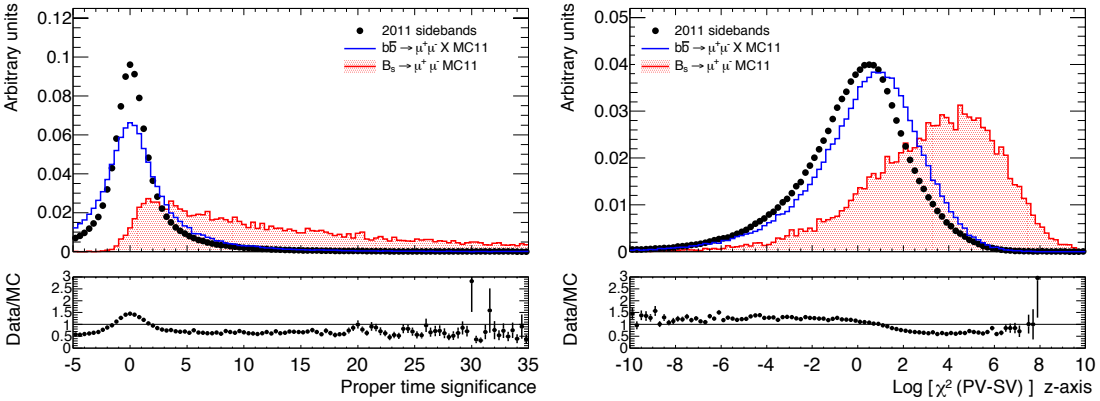


(c) *Re-weighting to the sidebands and a $L_{xy} > 0.2$ mm cut are applied.*

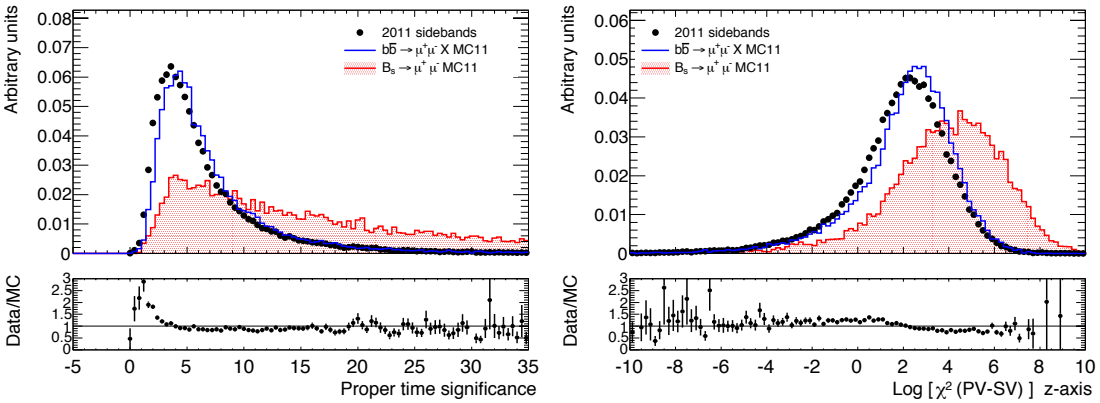
Fig. 5.16. *Distributions of ΔR (left), and L_{xy} (right). The sideband data (black dots) are compared with the $b\bar{b}$ background MC (blue solid histogram). The signal is shown for shape comparison (red-filled solid histogram). For each plot it is also reported the ratio between data and MC distributions.*



(a) *No re-weighting is applied.*

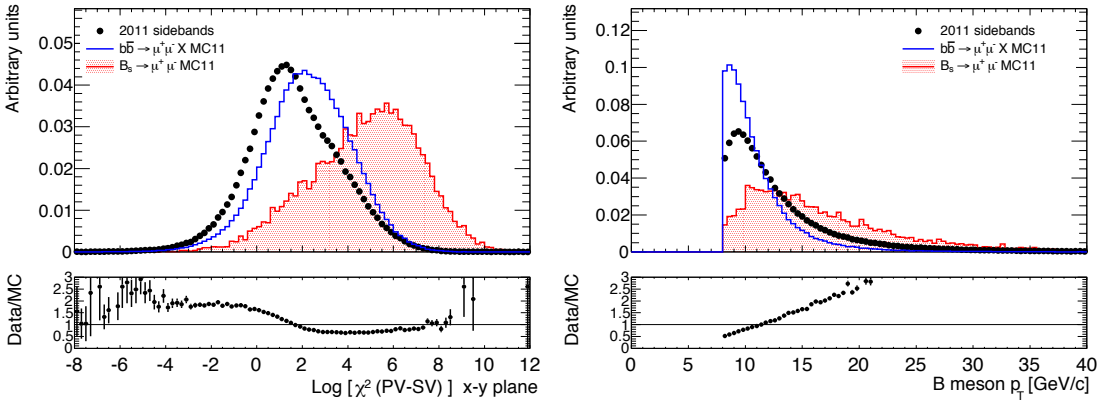


(b) *Re-weighting to the sidebands p_T and η distributions is applied.*

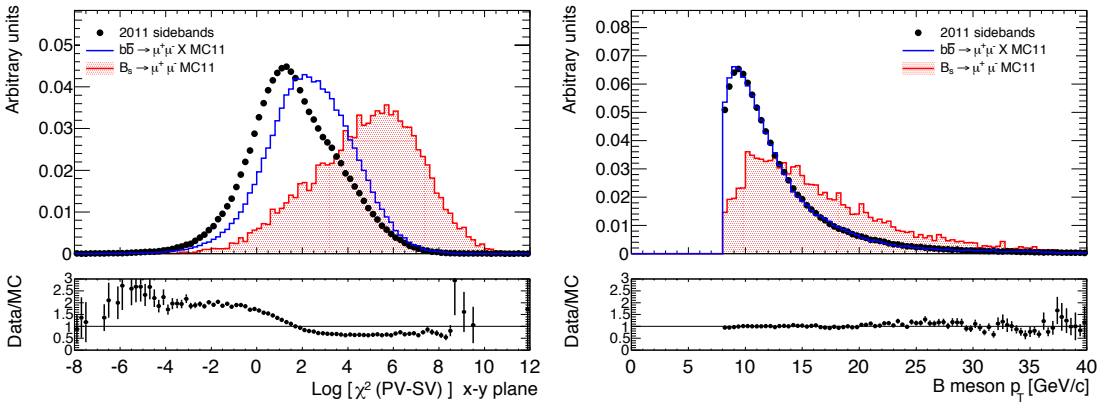


(c) *Re-weighting to the sidebands and a $L_{xy} > 0.2$ mm cut are applied.*

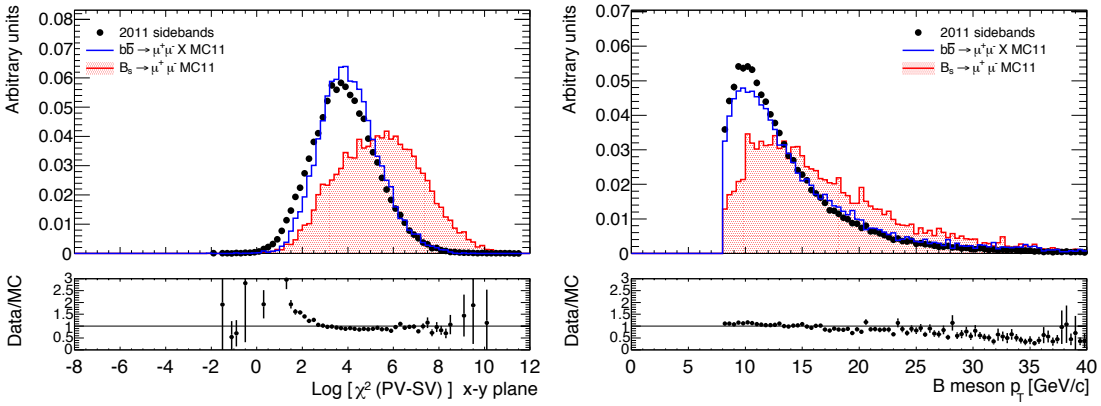
Fig. 5.17. *Distributions of proper time significance (left), and χ^2_{1D} (right). The sideband data (black dots) are compared with the $b\bar{b}$ background MC (blue solid histogram). The signal is shown for shape comparison (red-filled solid histogram). For each plot it is also reported the ratio between data and MC distributions.*



(a) *No re-weighting is applied.*

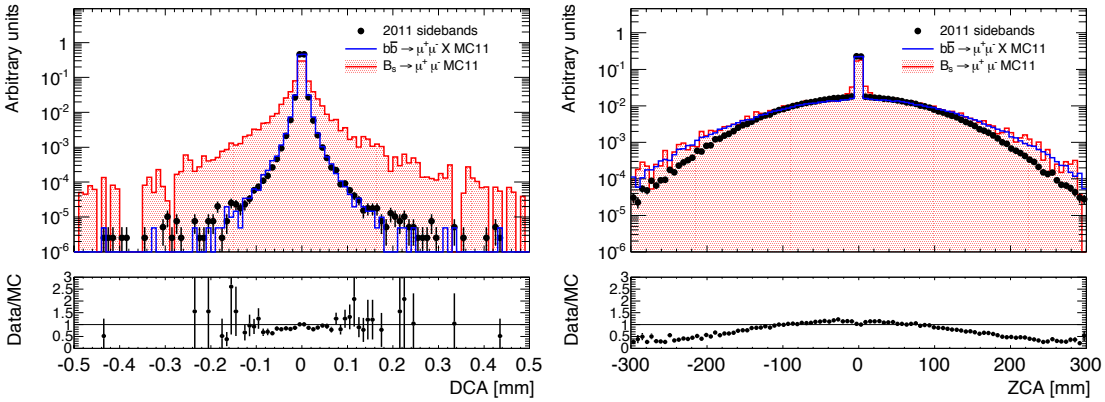


(b) *Re-weighting to the sidebands p_T and η distributions is applied.*

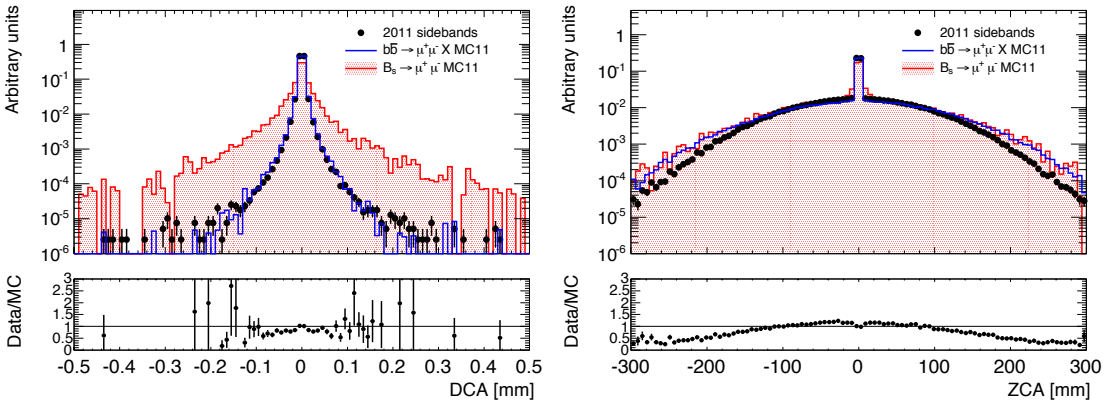


(c) *Re-weighting to the sidebands and a $L_{xy} > 0.2$ mm cut are applied.*

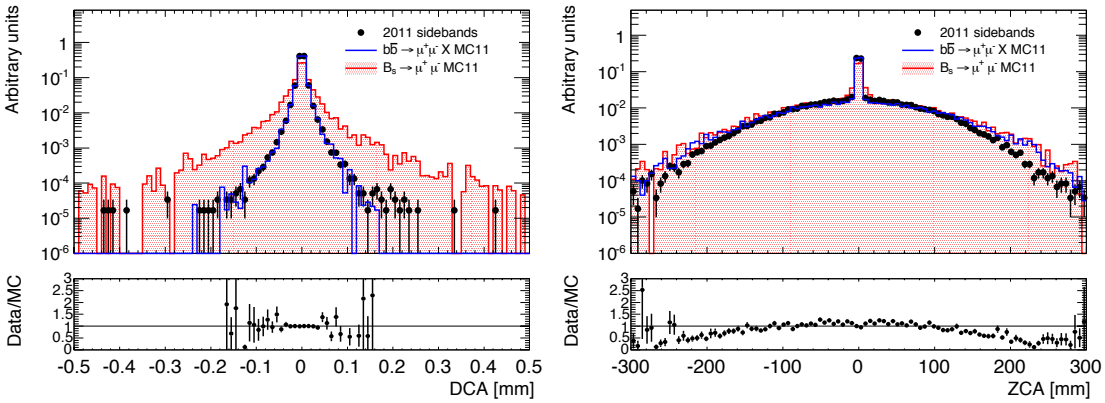
Fig. 5.18. *Distributions of χ^2_{2D} (left) and B transverse momentum, p_T (right). The sideband data (black dots) are compared with the $b\bar{b}$ background MC (blue solid histogram). The signal is shown for shape comparison (red-filled solid histogram). For each plot it is also reported the ratio between data and MC distributions.*



(a) *No re-weighting is applied.*

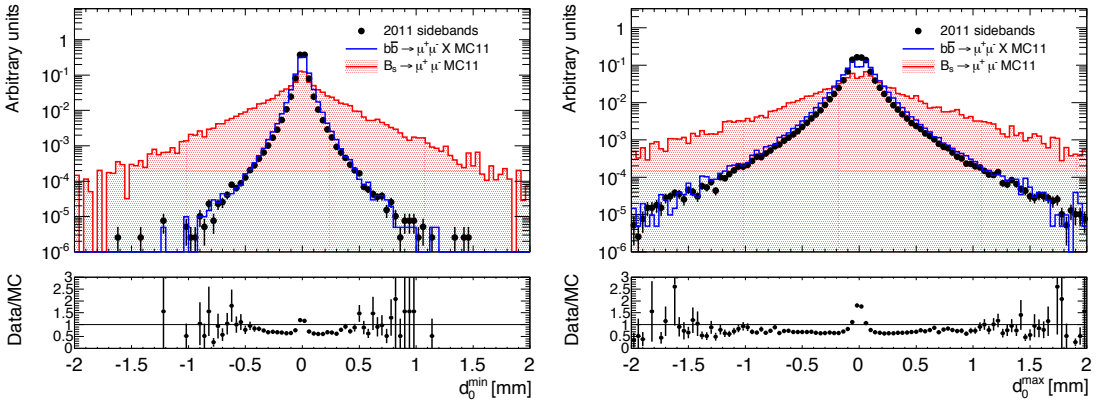


(b) *Re-weighting to the sidebands p_T and η distributions is applied.*

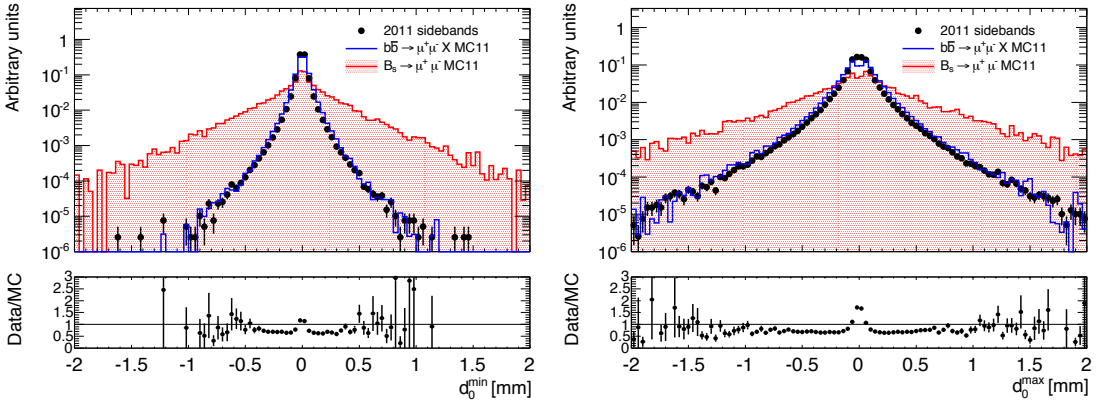


(c) *Re-weighting to the sidebands and a $L_{xy} > 0.2$ mm cut are applied.*

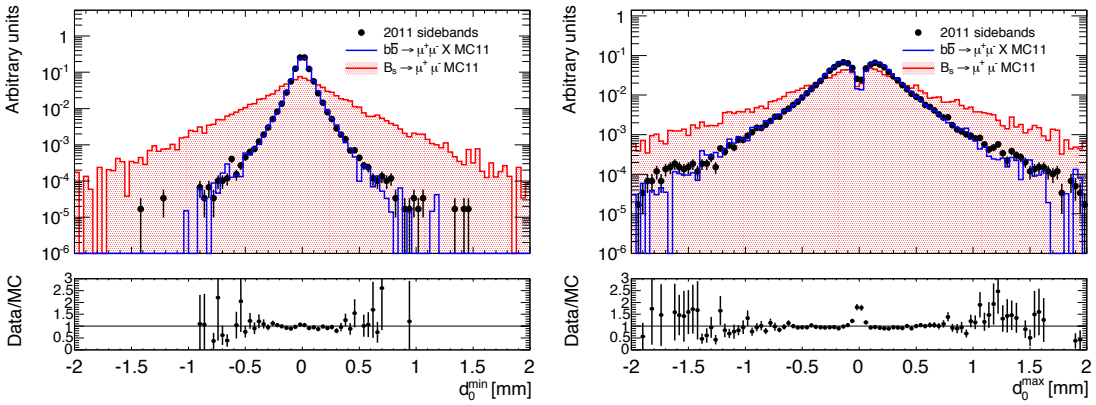
Fig. 5.19. *Distributions of Distance of Closest Approach (DCA) (left) and Distance of Closest Approach in z direction (ZCA) (right). The sideband data (black dots) are compared with the $b\bar{b}$ background MC (blue solid histogram). The signal is shown for shape comparison (red-filled solid histogram). For each plot it is also reported the ratio between data and MC distributions.*



(a) *No re-weighting is applied.*

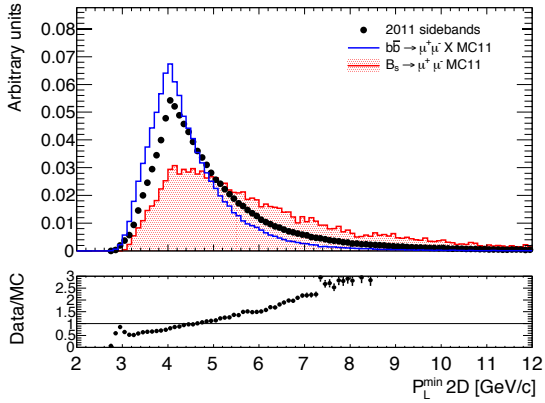


(b) *Re-weighting to the sidebands p_T and η distributions is applied.*

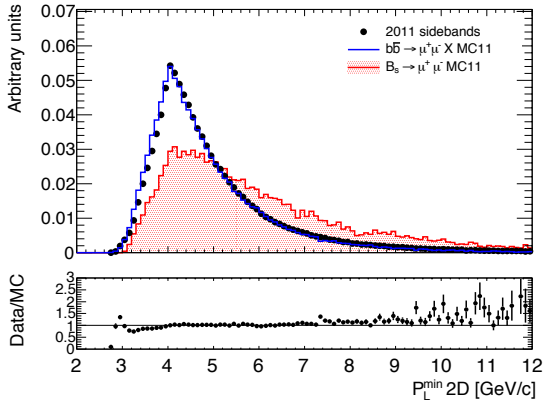


(c) *Re-weighting to the sidebands and a $L_{xy} > 0.2$ mm cut are applied.*

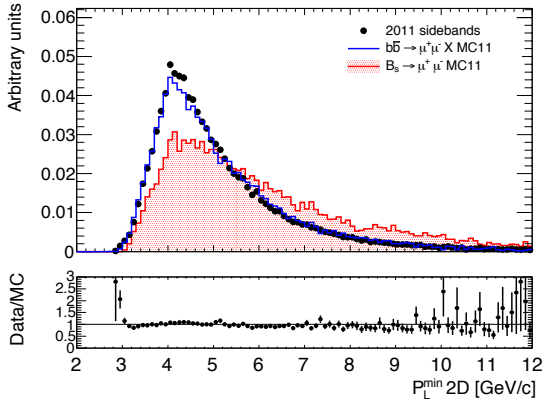
Fig. 5.20. *Distributions of impact parameter d_0^{min} (left) and impact parameter d_0^{max} (right). The sideband data (black dots) are compared with the $b\bar{b}$ background MC (blue solid histogram). The signal is shown for shape comparison (red-filled solid histogram). For each plot it is also reported the ratio between data and MC distributions.*



(a) *No re-weighting is applied.*



(b) *Re-weighting to the sidebands p_T and η distributions is applied.*



(c) *Re-weighting to the sidebands and a $L_{xy} > 0.2$ mm cut are applied.*

Fig. 5.21. *Distributions of P^{min}_{lng}. The sideband data (black dots) are compared with the $b\bar{b}$ background MC (blue solid histogram). The signal is shown for shape comparison (red-filled solid histogram). For each plot it is also reported the ratio between data and MC distributions.*

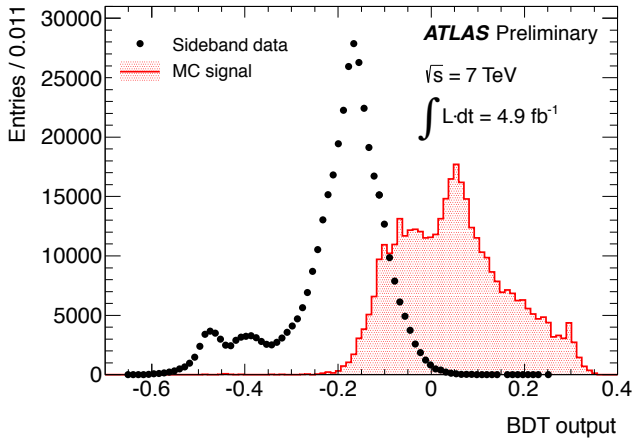


Fig. 5.22. *BDT distributions for sideband data (black dots) and signal (red-filled solid histogram). The areas are normalised to the number of entries in the sideband data.*

5.7 Yield Extraction for the Reference Channel

The reference channel yield has been extracted using an un-binned maximum likelihood fit to the distribution of the invariant mass $m_{\mu^+\mu^-K^\pm}$ of the $\mu^+\mu^-K^\pm$ system with a per-event mass error $\delta m_{\mu^+\mu^-K^\pm}$ (see figure 5.23).

Three different backgrounds contribute to the $B^\pm$ invariant mass window. To the left of the signal peak we can find the partially reconstructed B decays (e.g. $B^{+/0} \to K^{*+/0}$, $B^+ \to K^+\chi_{c1,c2}$) where one or more of the final state particles are missed in the reconstruction. Close (on the right) to the signal peak we have the reflection of the Cabibbo suppressed $B^\pm \to J/\psi\,\pi^\pm$ decay in which the kaon mass is associated to the pion, shifting the peak to higher masses of roughly the difference between K and π mass (the fraction of $B^\pm \to J/\psi\,\pi^\pm$ over $B^\pm \to J/\psi\,K^\pm$ is about 4.8 % [121]). Finally, the combinatorial background (mostly composed, after the analysis selection, by $b\bar{b} \to J/\psi\,X$) that smoothly crosses the mass range and that is greatly reduced after having applied the BDT optimised selection. To avoid any bias from the DD re-weighting procedure of the Monte-Carlo samples (described in section 5.3), only even-numbered events are used in the extraction of the reference channel yield.

The signal $B^\pm \to J/\psi\,K^\pm$ has been modelled using a single Gaussian with $\sigma = s{\cdot}\delta m$

$$\mathcal{M}_s^m(m|\mu_s^m, s, \delta m) = e^{-\frac{(m-\mu_s^m)^2}{2\,(s\cdot\delta m)^2}} \tag{5.9}$$

where δm is the event-by-event uncertainty (taken from the MC), propagated from the uncertainty on the B-candidate vertex fit. The normalisation, the mean and the scaling parameter s are left floating in the fit. The mass error has been modelled from the signal MC. The combinatorial has been shaped using an exponential function

$$\mathcal{M}_{comb}^m(m|\alpha_{comb}) = e^{\alpha_{comb}\cdot m} \tag{5.10}$$

with the mass error parameterisation taken from the odd-numbered left and right sidebands in data. The normalisation and the shape are extracted from the fit on data. The partially reconstructed B decays have been shaped using a complementary error function [122]

$$\mathcal{M}^m_{p.r.}(m \,|\, \mu^m_{p.r.}, \sigma^m_{p.r.}) = \mathrm{Erfc}\left(\frac{m - \mu_{p.r.}}{\sigma_{p.r.}}\right) \tag{5.11}$$

with shape parameters and normalisation free to float in the fit. The error function for these contributions has been extracted from odd-numbered left and right sidebands in data. Finally, the shape of $B^\pm \to J/\psi\,\pi^\pm$ decays is modelled using a Crystal Ball function

$$\mathcal{M}^m_{J/\psi\pi}(m \,|\, \mu^m_{J/\psi\pi}, \sigma^m_{J/\psi\pi}, \alpha^m_{J/\psi\pi}, k^m_{J/\psi\pi}) =$$
$$\begin{cases} e^{-t^2/2} & \text{if } t \geqslant -|\alpha^m_{J/\psi\pi}| \\ \left(\frac{k}{|\alpha^m_{J/\psi\pi}|}\right)^k \cdot e^{-|\alpha^m_{J/\psi\pi}|^2/2} \cdot \left(\frac{k}{|\alpha^m_{J/\psi\pi}|} - |\alpha^m_{J/\psi\pi}| - t\right)^{-k} & \text{if } t < -|\alpha^m_{J/\psi\pi}| \end{cases} \tag{5.12}$$

with $t = (m - \mu^m_{J/\psi\pi})/\sigma^m_{J/\psi\pi}$, where $\mu^m_{J/\psi\pi} = 5353$ MeV, $\sigma^m_{J/\psi\pi} = 53.8$ MeV, $\alpha^m_{J/\psi\pi} = -1.28$ and $k = 145$ are the values obtained from the fit to the MC. The normalisation is left floating in the fit and the mass error has been modelled from the MC $J/\psi\pi^\pm$. All the distributions of the event-by-event mass error are modelled using a non parametric kernel estimation RooKeyPdf [120] with smoothing factor $\rho = 2$.

The result of the fit is reported in figure 5.23. The mass spectrum of the $B^\pm$ (left) and the mass error distribution (right), obtained from even-numbered events passing all selection cuts, are shown along with the un-binned maximum likelihood fit overlaid on them.

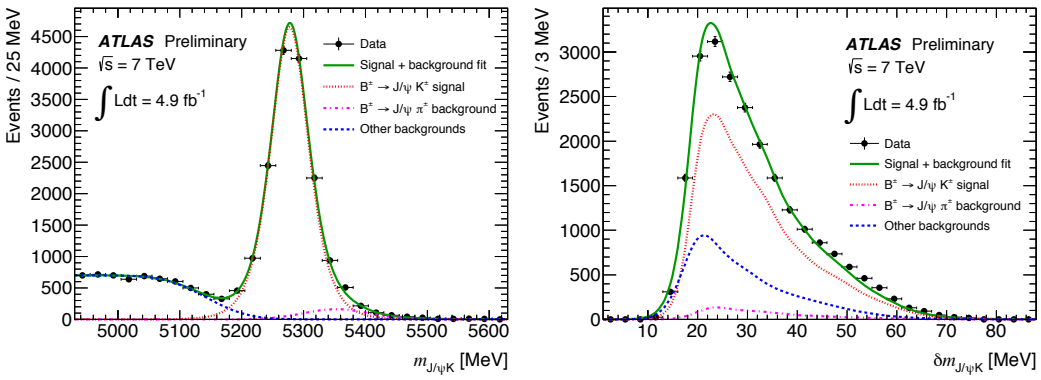


Fig. 5.23. *$B^\pm$ invariant mass (left) and mass error (right) distributions with the corresponding un-binned maximum likelihood fit overlaid. The black dots are data, the solid green line is the total fit projection. The dotted red line is the $B^\pm \to J/\psi\,K^\pm$ signal, the dash-dotted magenta curve is the background from $B^\pm \to J/\psi\,\pi^\pm$, and the dashed blue curve corresponds to the sum of the partially reconstructed B decays and the combinatorial background components.*

The systematic uncertainties on the fit are estimated using two other background

models: first without the $B^{\pm} \to J/\psi\,\pi^{\pm}$ Crystal Ball function, and then using a linear function for the combinatorial background instead of an exponential.

A second fit strategy, based on a two-dimensional un-binned extended maximum likelihood fit to data with simultaneous inclusion of three MC models for several fit components, has been developed in parallel to the one reported above. Both implementations provide a consistent result for the reference channel yield. The combined measurement gives 15,214 even-numbered events with an uncertainty of $\pm\,1.1\,\%$ (stat.) and $\pm\,2.4\,\%$ (syst.). For the combination of the two main results the weighted average of their central values (taking the inverse of the square of their statistical uncertainties as weights) has been computed. Similarly, an average statistical uncertainty on the combined central values, since the un-binned maximum likelihood fits have been performed on the identical data sample, has been calculated. For the determination of the systematic uncertainties the differences of the central values of all models (the default ones of the two fits as well as the ones with alternative background models) to the central value derived by the above averaging procedure have been considered. The maximum difference is assigned as a symmetric estimate for the systematic uncertainty equal to $2.4\,\%$ and $0.3\,\%$ for the default and the alternative fit strategy respectively. This conservative approach does not harm the sensitivity of the $B_s^0 \to \mu^+\mu^-$ measurement since it has been proved that a factor two applied to the size of the systematic uncertainties assigned to the $B^{\pm}$ yield numbers has no effect on the expected upper limit because other uncertainties (summarised in section 5.9) dominate.

5.8 Acceptance times Efficiency Ratio ($A \times \varepsilon$)

The ratio of the acceptance A and selection efficiency ε terms for the $B_s^0 \to \mu^+\mu^-$ signal and the $B^{\pm} \to J/\psi\,K^{\pm}$ reference channel, $R_{A\varepsilon} = \frac{A_{J/\psi\,K^{\pm}}}{A_{\mu^+\mu^-}} \cdot \frac{\varepsilon_{J/\psi\,K^{\pm}}}{\varepsilon_{\mu^+\mu^-}}$, is evaluated using the MC after having applied both the GL corrections and DD weights. The term $A \times \varepsilon$ corresponds to the number of events passing the baseline and optimised selection, normalised to the phase-space volume defined by $p_T^B > 8$ GeV and $|\eta^B| < 2.5$, and for the 2011 analysis was equal to 0.267 ± 1.8 % (stat.) $\pm\,6.9$ % (syst.).

The systematic uncertainties are associated with the GL corrections and DD weights applied to the MC samples, and some residual discrepancies between data and MC. The impact of these sources has been evaluated using toy MC experiments. The GL and DD corrections have been varied within their uncertainties and the ratio $R_{A\varepsilon}$ has been reevaluated for each toy experiment. The systematic uncertainty on $R_{A\varepsilon}$ due to the residual discrepancies between data and MC distributions has been estimated by folding the observed data-MC discrepancies on the $B^{\pm} \to J/\psi\,K^{\pm}$ channel as additional weights into the MC samples and assigning the change in the value of $R_{A\varepsilon}$ as systematic uncertainty. The change in $A \times \varepsilon$ is highly correlated between the signal and the reference channels and almost cancels out in the ratio. This is true for all but the isolation variable. The latter has been considered separately, as it depends on the B flavour produced and so separate evaluations have to be performed for the

$B^{\pm}$ of the reference channel and the B_s^0 of the signal. For this reason, the estimate of the data-MC discrepancy on the signal for the isolation variable is taken from the $B_s^0 \to J/\psi\phi$ control sample. The total systematic uncertainty on $R_{A\varepsilon}$ amounts to $\Delta R_{A\varepsilon}/R_{A\varepsilon} = \pm 6.9\,\%$ while the statistical uncertainty due to the finite size of the MC sample is $\pm 1.8\,\%$.

5.9 Systematic uncertainties

As already discussed in the previous section, due to the relative measurement of the $B_s^0 \to \mu^+\mu^-$ branching ratio with respect to the reference channel, the systematic uncertainty in $A \times \varepsilon$ almost cancels out in the ratio. The ingredients in the branching ratio evaluation that are affected by systematic uncertainties are the SES (equation 5.4) and the number of background events in the signal region. The different sources of systematic are reported in table 5.6. These come from the product of the branching ratios with the relative b hadronisation probability, the $A \times \varepsilon$ ratio $R_{A\varepsilon}$ (dominated by the systematic uncertainties introduced by the residual data-MC discrepancies), $B^{\pm}$ yield, $B_s^0/B^{\pm}$ relative vertexing efficiency, and $K^{\pm}$ tracking efficiency. In addition, there is a systematic introduced by the $K^{\pm}$ charge dependent reconstruction asymmetry, due to the use of the $B^+ \to J/\psi\, K^+$ MC for the efficiency evaluation.

Systematic source	Contribution
PDG branching ratio and f_s/f_d	8.5 %
$K^{\pm}$ tracking efficiency	5 %
Vertexing efficiency	2 %
$K^{\pm}$ charge asymmetry in $B^{\pm} \to J/\psi K^{\pm}$	1 %
$B^{\pm} \to J/\psi\, K^{\pm}$ yield	2.4 %
$R_{\varepsilon A}$	6.9 %
Total (combined in quadrature)	12.5 %

Table 5.6. *Summary of the $\Delta SES/SES$ error due to the sources of systematic uncertainty on the single event sensitivity.*

The background interpolation factor R_{bkg}^{obs}, defined as half of the ratio of the combined widths of the sideband regions (600/2 MeV) and the width of the signal region $2 \cdot \Delta m$ (since only half of the sidebands is used for the interpolation, while the background contribution has to be estimated to all the events in the signal region), is an effective mass width ratio with respect to the sidebands. The systematic uncertainty on R_{bkg}^{obs} (evaluated by comparing linear, exponential and exponential+linear interpolations) accounts for the effect of the mass dependence of the continuum background and for the additional background components in the low mass sideband, and has been estimated to be 4 % [107]. The systematic related to the uncertainty in the estimation of the number $N_{hh'}$ is negligible due to the very small contribution of the resonant $B \to hh'$ background in the signal region.

Quantity	Value
$N_{J/\psi K^\pm}$	$15{,}214 \pm 1.10\,\% \pm 2.39\,\%$
$R_{\varepsilon A}$	$0.267 \pm 1.8\,\% \pm 6.9\,\%$
SES	$(2.07 \pm 0.26) \cdot 10^{-9}$
R^{obs}_{bkg}	1.240 ± 0.050
$N^{obs}_{bkg,SB}$	8
$N_{B\to hh'}$	0.30
$N^{exp}_{SR} \mid N^{obs}_{SR}$	6.75 | 6

Table 5.7. *Input values used for the extraction of the upper limit using the CL_s method. The first two rows show the $B^\pm$ yield $N_{J/\psi K^\pm}$ and the ratio of the selection efficiency times acceptance for the B^0_s signal and the $B^\pm$ reference channel, $R_{\varepsilon A}$, used in the evaluation of the SES.*

5.10 Branching Ratio Upper Limit Extraction

The upper limit on the $B^0_s \to \mu^+\mu^-$ branching ratio is evaluated following the ATLAS prescription for the extraction of frequentist limits by means of the standard implementation [124, 125] of the $\mathrm{CL_s}$ method [126] (the $\mathrm{CL_s}$ approach will be reviewed in more detail in chapter 6). The likelihood used is the following[10]

$$\begin{aligned}\mathcal{L} \;=\; & \mathrm{Poisson}(N^{obs}_{SR}|\varepsilon \cdot \mathrm{BR} + N_{bkg} + N_{B\to hh'}) \cdot \mathrm{Poisson}(N^{obs}_{bkg,SB}|R_{bkg}\, N_{bkg}) \\ & \times \mathrm{Gauss}(\varepsilon^{obs}|\varepsilon, \sigma_\varepsilon)\, \mathrm{Gauss}(R^{obs}_{bkg}|R_{bkg}, \sigma_{R_{bkg}}) \end{aligned} \tag{5.13}$$

where BR is the branching ratio under study, ε the inverse of the SES and N_{bkg} the number of continuum background in the signal region ($N^{obs}_{bkg,SB}/R_{bkg}$). The two directly measured quantities are the number of total observed events in the signal region (N^{obs}_{SR}) and the even-numbered observed events in the two sidebands ($N^{obs}_{bkg,SB}$). The number of expected background events in the signal region (N^{exp}_{SR}) is obtained summing the $N_{B\to hh'}$ contribution to the background interpolated from the sidebands ($N^{obs}_{bkg,SB}/R^{obs}_{bkg}$). R_{bkg} and ε are treated as nuisance parameters which are constrained to the corresponding observables R^{obs}_{bkg} and ε^{obs} within their uncertainties by the Gaussian terms. Table 5.7 summarises all the inputs necessary for the evaluation of the $B^0_s \to \mu^+\mu^-$ branching ratio.

Before the un-blinding, the median of the expected upper limit, obtained setting the count in the signal region (N^{obs}_{SR}) equal to N^{exp}_{SR} that is the background interpolated from the sidebands plus the small contribution from the resonant background, was $1.6^{+0.7}_{-0.4} \cdot 10^{-8}$ ($1.3^{+0.6}_{-0.4} \cdot 10^{-8}$) at 95% CL (90% CL), where the range encloses 68 % ($\pm 1\,\sigma$) of the background-only pseudo-experiments.

[10] In this notation, for a Poisson distribution $\mathrm{Poisson}(N|\mu)$, N is the event count and μ is the mean, while for a Gaussian distribution $\mathrm{Gauss}(V|\nu,\sigma)$, V is the observed value, μ and σ are the mean and the standard deviation respectively.

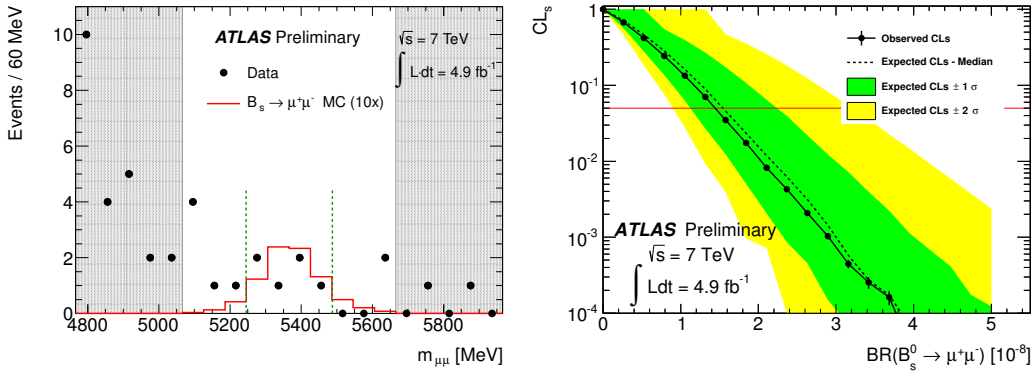


Fig. 5.24. *Left: invariant-mass distribution of selected candidates in data (dots). The plot also indicates the signal (continuous line) as predicted by MC assuming* $BR(B_s^0 \to \mu^+\mu^-) = 3.5 \cdot 10^{-8}$ *[127] (increased by a factor 10), the optimised signal window (two dashed vertical lines) and the sidebands used in the analysis (grey areas). The expected number of* $B_s^0 \to \mu^+\mu^-$ *signal in the signal region is* 1.7 ± 0.2 *events. The expected background yield per bin in the signal region is 1.7 events. Right: observed* CL_s *(circles) as a function of* $BR(B_s^0 \to \mu^+\mu^-)$*. The 95% CL limit is indicated by the horizontal (red) line. The dark (green) and light (yellow) bands correspond to the* $\pm 1\,\sigma$ *and* $\pm 2\,\sigma$ *ranges of the background-only pseudo-experiments with the median of the expected* CL_s *given by the dashed line.*

After the un-blinding, 6 events have been observed in the signal region (N_{SR}^{obs}) giving an observed upper limit $\mathrm{BR}(B_s^0 \to \mu^+\mu^-) < 1.5\,(1.2) \cdot 10^{-8}$ at 95% (90%) CL. Figure 5.24 shows the distribution of the observed $\mu^+\mu^-$ candidates and the $\mathrm{CL_s}$ behaviour for different tested values of the $\mathrm{BR}(B_s^0 \to \mu^+\mu^-)$, evaluated using 300,000 toy Monte Carlo simulations per point.

Chapter 6

Additional Studies on 2011 Data

The analysis performed on the full 2011 dataset, reported in chapter 5, was essentially based on the same strategy adopted for the published one on 2.4 fb^{-1} of data, with the introduction of some novelties, like the BDT training performed on MC background sample.

The 8 TeV dataset provides more than four times the statistic used for the previous analysis and it can give the opportunity to measure the $B_s^0 \to \mu^+\mu^-$ SM branching ratio and set an upper limit on the $B^0 \to \mu^+\mu^-$ decay. Unfortunately, as it will be shown in chapter 7, ATLAS, compared with the CMS and LHCb experiments, has limited trigger efficiency and mass resolution, resulting in a degraded sensitivity to these decays. To partially fill these gaps, efforts should be committed to optimise the analysis, with particular attention to the selection strategies and the signal extraction procedure. For this purpose, the 2011 un-blinded dataset represents an extremely precious sample on which the new strategies can be tested.

In this chapter are documented the studies I performed on 2011 data and MC, aiming to improve the selection procedure for the 2012 analysis. I will report the tests on the discriminating variables, trying also to better understand all the possible background contributions (section 6.1), and the ones on the invariant mass fit for the signal yield extraction (section 6.2). The reliability of the fit and its performance will be tested using both toy MC experiments (section 6.2.4) and 2011 un-blinded data extracting the upper limit on the $B_s^0 \to \mu^+\mu^-$ BR using the CL_s method (section 6.3).

All these tests are very preliminary and they have been performed when the analysis on 2011 data was still in progress. Therefore, some results have been obtained with a preliminary version of 2011 MC and data, while some others using the final selection that has been adopted for the public 2011 analysis. For "didactic purposes" the code used to implement the invariant mass fit, the toy MC experiments and the CL_s was developed from scratch.

What is reported in this chapter is not documented in the ATLAS internal note for the 2012 analysis.

6.1 Additional Studies on the Background Rejection

For the 2011 analysis, the dependence of the multivariate selection on the di-muon mass has been checked studying the linear correlation of the discriminating variables with the mass that showed to be negligible. Nonetheless, further investigations are mandatory to be able to reliably perform an invariant mass fit.

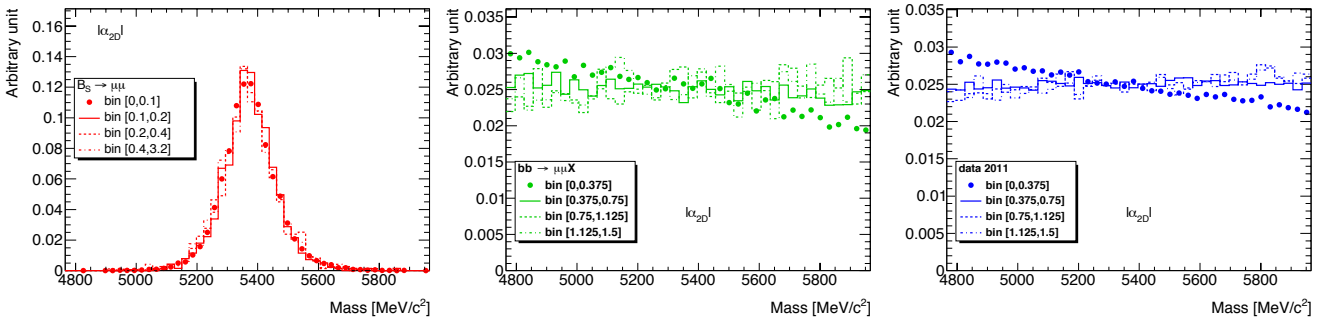
6.1.1 Di-muon Mass Distribution for Different Bins of the Discriminating Variables

The distribution of each of the 13 discriminating variables used into the 2011 analysis has been subdivided into four bins, and the invariant mass spectrum of the events contained in each of them is shown. Figures 6.1, 6.2, 6.3 and 6.4 report the di-muon mass for $B_s^0 \to \mu^+\mu^-$ signal, $b\bar{b} \to \mu^+\mu^- X$ background and 2011 un-blinded data in the four bins, while table 6.1 shows the values for the linear correlation between the mass and the discriminating variables, evaluated for each of the above datasets (the corresponding scatter plots are reported in appendix B.1).

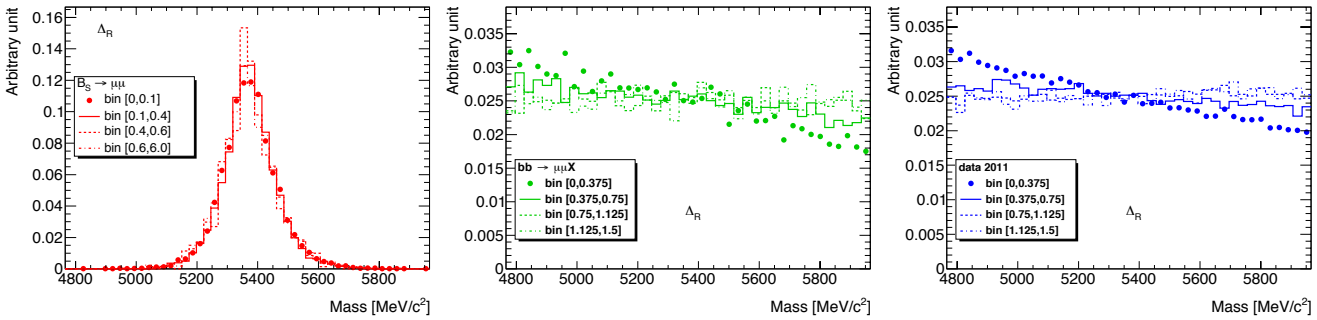
Discriminating variable	**Di-muon mass linear correlation [%]**		
	Signal	**MC background**	**2011 un-blinded data**
α_{2D}	- 1.8	1.3	0.7
ΔR	- 1.2	1.5	1.3
Isolation ($I_{0.7}$)	1.8	2.9	1.6
$\log(\chi^2$ (PV-SV)) z-axis	1.9	- 1.4	- 0.5
$\log(\chi^2$ (PV-SV)) in x-y	0.5	- 0.6	0.4
L_{xy}	- 0.9	- 3.2	- 3.2
Proper time significance	- 0.7	- 1.5	- 0.9
d_0^{max}	0.2	- 0.1	0.1
d_0^{min}	0.1	- 0.1	0.1
ZCA	- 0.6	- 0.4	- 0.2
DCA	0.6	0.2	0.1
B meson p_T	3.9	- 1.7	- 2.0
p_L^{min}	2.7	- 5.0	- 4.2

Table 6.1. *Values for the linear correlations between the mass and the discriminating variables for $B_s^0 \to \mu^+\mu^-$ signal, $b\bar{b} \to \mu^+\mu^- X$ MC background and 2011 un-blinded data.*

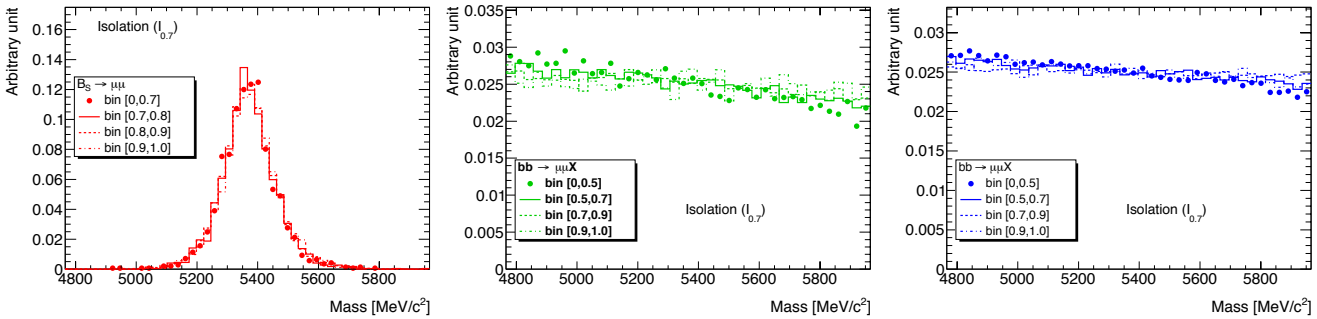
A non-zero correlation with the mass of any of the discriminating variables can reflect on a dependence of the BDT output on the B_s^0 mass. Figure 6.5 reports the invariant mass distribution for 2011 un-blinded data and $b\bar{b} \to \mu^+\mu^- X$ background MC in eight bins of the BDT output (the bins with higher values correspond to the signal-like region of the BDT spectrum). The higher the value of the classifier is, the higher seems to be the mass dependence (the same behaviour is not so evident in the



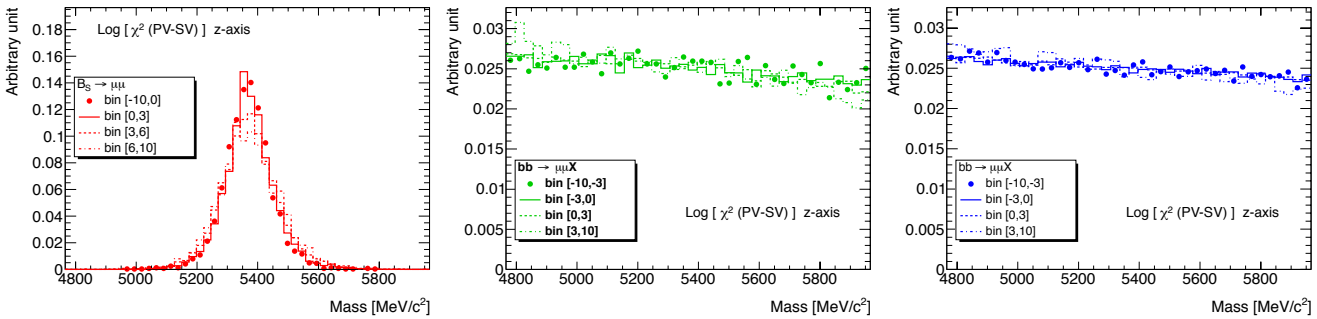
(a) *Pointing angle* $|\alpha_{2D}|$



(b) ΔR

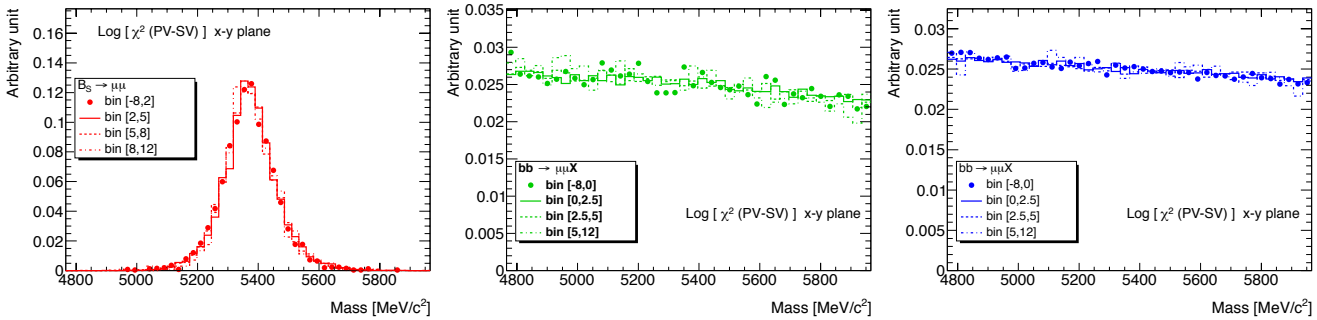


(c) *Isolation* $(I_{0.7})$

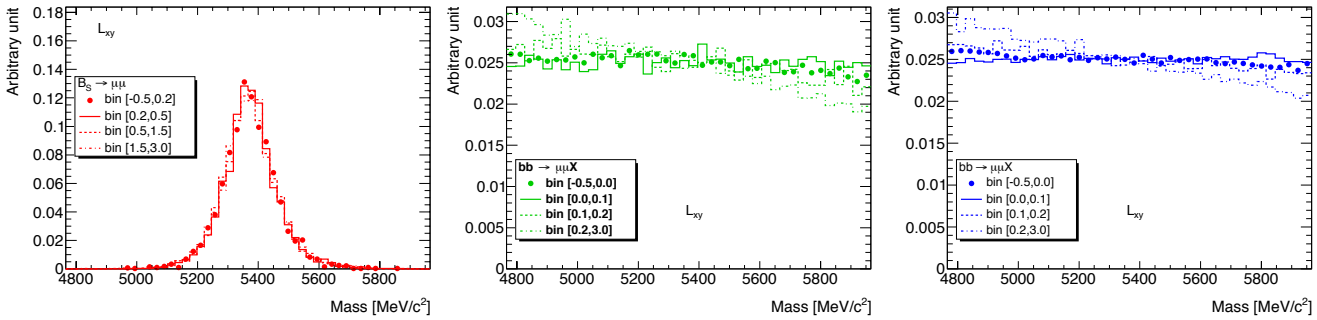


(d) $\log(\chi^2(PV-SV))$ *along z-axis*

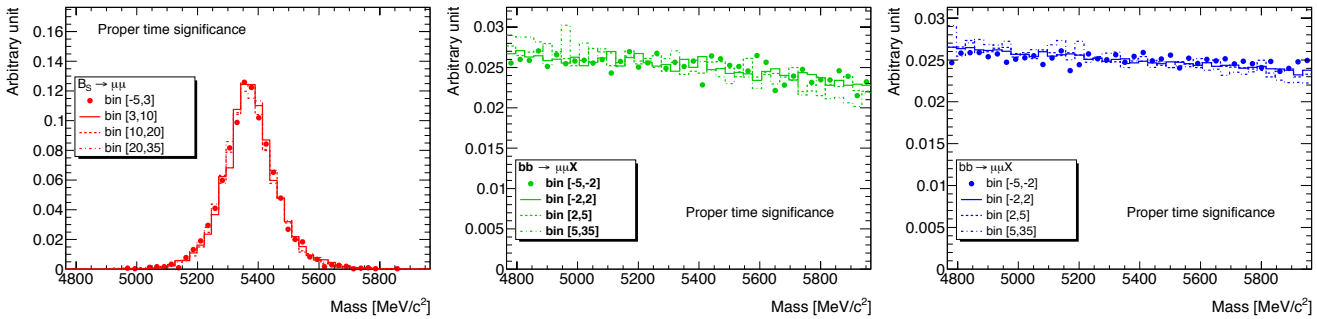
Fig. 6.1. *Invariant mass distribution for* $B_s^0 \to \mu^+\mu^-$ *signal (left),* $b\bar{b} \to \mu^+\mu^- X$ *background (middle) and 2011 un-blinded data (right) in four bins of the discriminating variables* α_{2D}*,* ΔR*, Isolation and* $\chi^2(PV-SV)$ *along z-axis.*



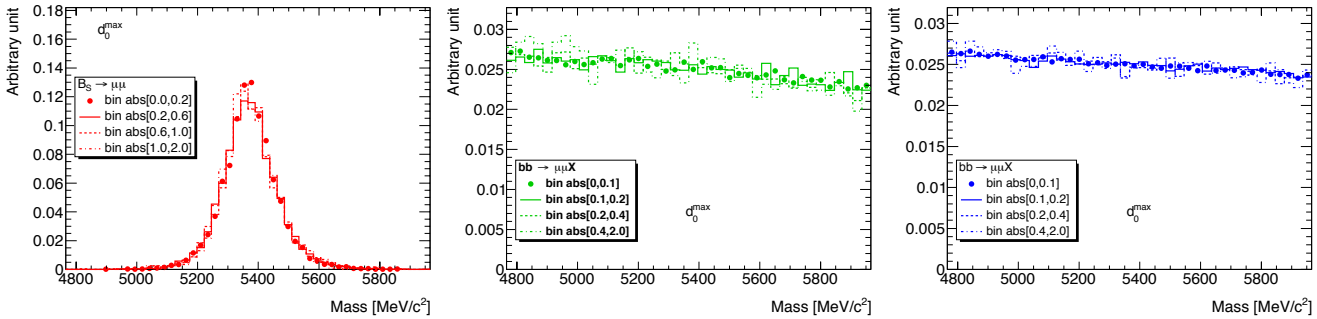
(a) $\log(\chi^2(PV-SV))$ *in the x-y plane*



(b) L_{xy}

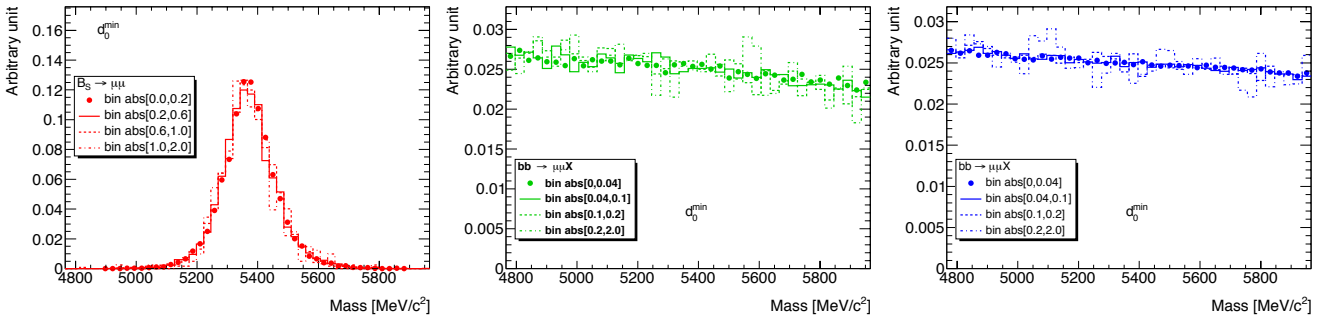


(c) *Proper time significance*

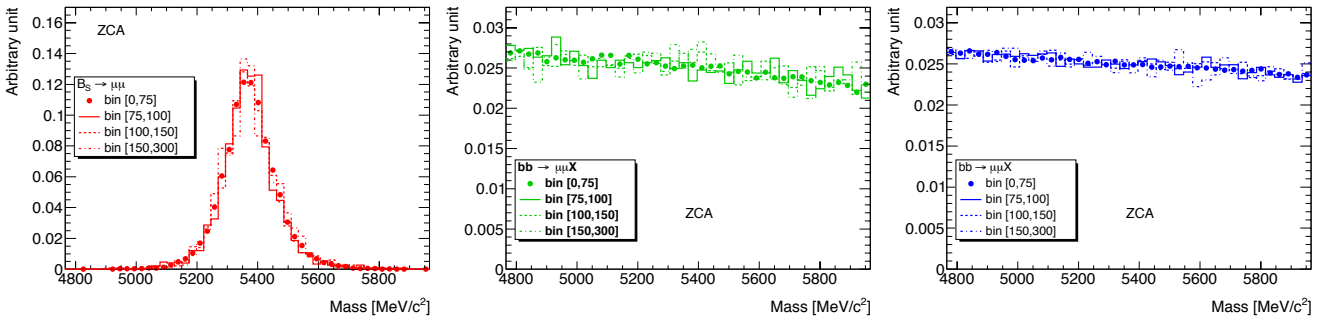


(d) d_0^{max}

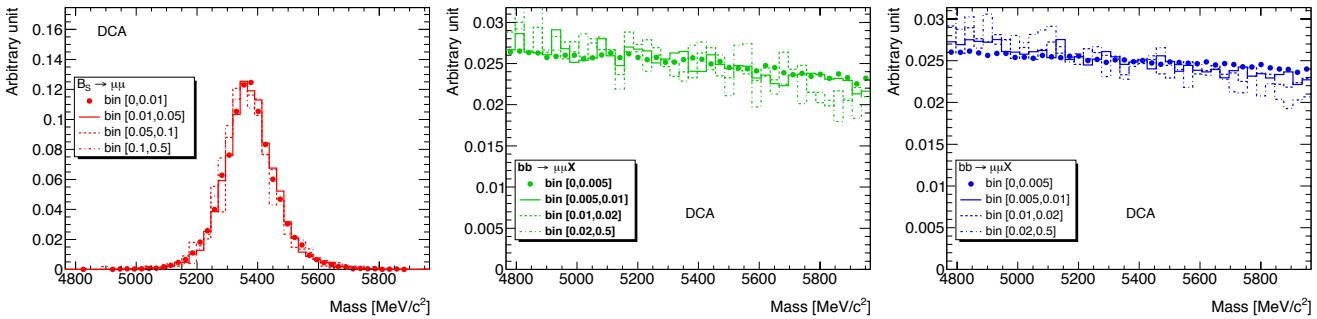
Fig. 6.2. *Invariant mass distribution for* $B_s^0 \to \mu^+\mu^-$ *signal (left),* $b\bar{b} \to \mu^+\mu^- X$ *background (middle) and 2011 un-blinded data (right) in four bins of the discriminating variables* $\chi^2(PV-SV)$ *in the x-y plane,* L_{xy}*, proper time significance and* d_0^{max}*.*



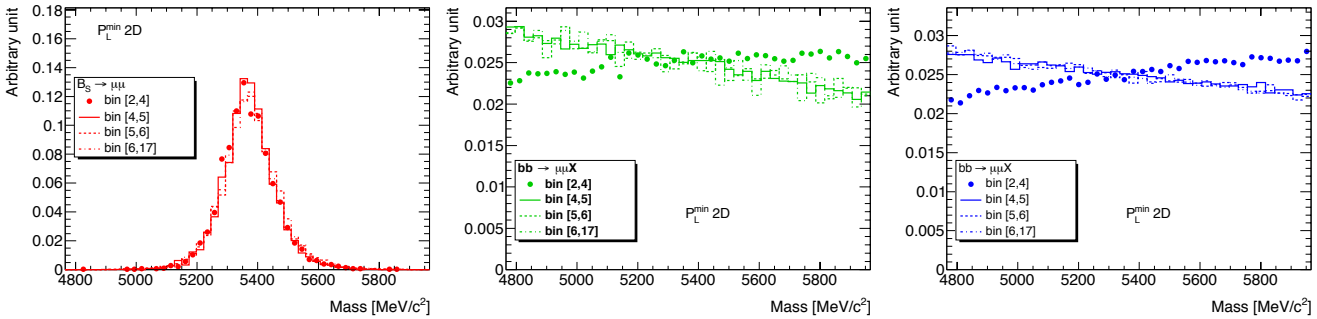
(a) d_0^{min}



(b) ZCA

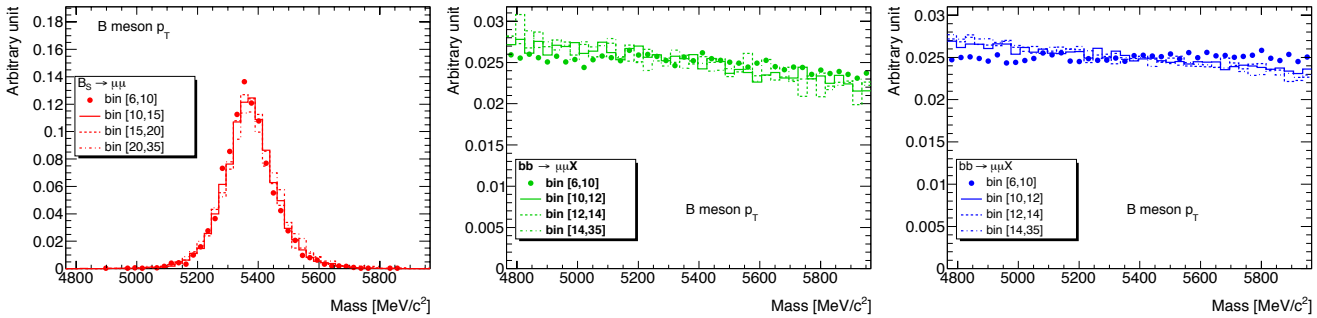


(c) DCA



(d) P_L^{min} $2D$

Fig. 6.3. *Invariant mass distribution for $B_s^0 \to \mu^+\mu^-$ signal (left), $b\bar{b} \to \mu^+\mu^- X$ background (middle) and 2011 un-blinded data (right) in four bins of the discriminating variables d_0^{min}, ZCA, DCA and p_L^{min}.*



(a) *B meson p_T*

Fig. 6.4. *Invariant mass distribution for $B_s^0 \to \mu^+\mu^-$ signal (left), $b\bar{b} \to \mu^+\mu^- X$ background (middle) and 2011 un-blinded data (right) in four p_T bins.*

MC sample). Admittedly, what figure 6.5 shows is not a strong BDT mass correlation, but rather the presence of background components different from the continuum.

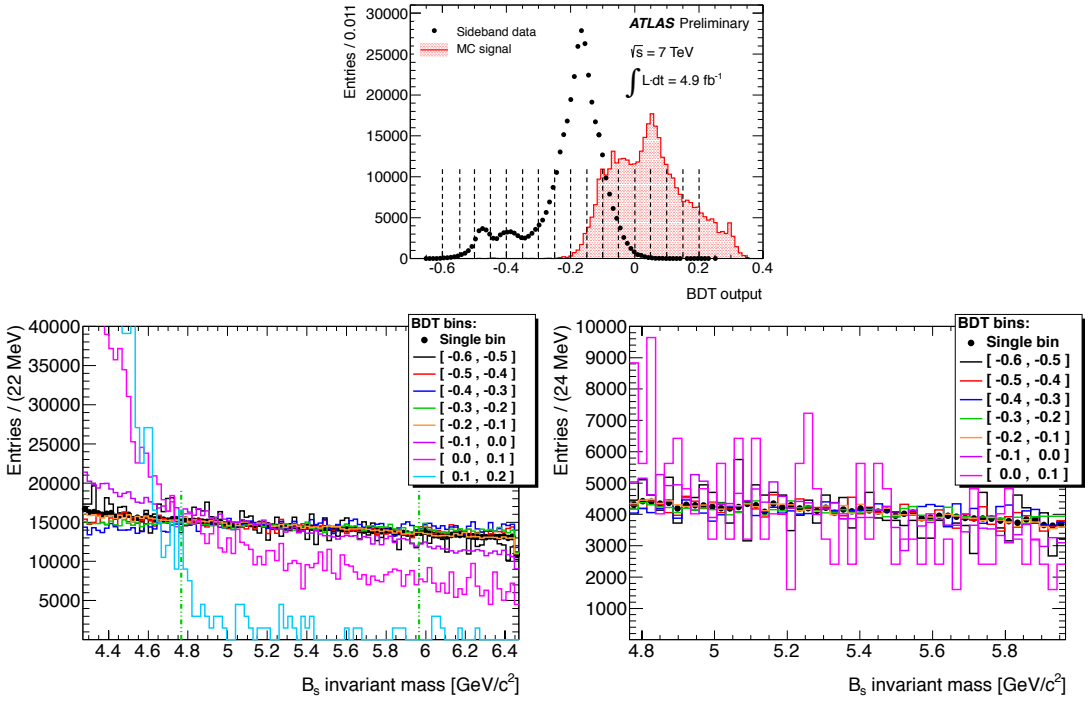


Fig. 6.5. Top*: BDT output distribution for 2011 sidebands (black dots) and signal $B_s^0 \to \mu^+\mu^-$ (red-filled solid histograms). The sidebands distribution has been divided in eights BDT bins (black vertical dashed lines).* Bottom*: invariant mass distribution obtained for the eight BDT bins for 2011 un-blinded data (left) and $b\bar{b} \to \mu^+\mu^- X$ MC background (right).*

The BDT has been trained and optimised to reject the combinatorial background and select the signal-like components present in data. As other experiments (like LHCb

and CMS) showed, the left sideband is clearly dominated by the continuum background, but also other components are present. In this part of the mass spectrum, background contributions coming from partially reconstructed and semi-leptonic decays may not be negligible. Since their topology is very close to that of the signal, their behaviour in terms of BDT output is mostly the same as $B_s^0 \to \mu^+\mu^-$. They therefore constitute an almost irreducible background. However, since the invariant mass of the two muons is peaked far below the B_s^0 mass, only the tail of the distribution might leak into the B_s^0 signal region.

This kind of background was already considered in the public 2011 analysis (see section 5.9 and figure 5.24) and it was taken into account as a systematic on the R_{bkg}^{obs} (4 %), but with the intent of performing an invariant mass fit for the signal extraction, this contribution should be reliably and separately modelled. Unfortunately, no MC was available in ATLAS at the time of these tests, but it has been produced for the 2012 analysis and more details will be reported in sections 8.2.1 and 8.2.2.

6.1.2 BDT Training on $B^\pm$ Right Sideband

As reported in chapter 5, for the 2011 analysis the BDT has been trained on MC background events, which have been re-weighted using sideband data to correct the inaccuracies present in the simulation. Ideally, the best solution would be to train the MVA directly on data. However, if the same data are used also for other purposes, like the estimation of the background in the signal region, this procedure could introduce a bias (even if small) in the final results. For that reason, the possibility of performing the training and testing procedure on an alternative sample has been explored.

The only dataset available for this purpose is the $B^\pm \to J/\psi\, K^\pm$. Although the event topology is different with respect to the B_s^0 signal (3-body versus 2-body decay), the same discriminating variables used to reject the combinatorial background could be very similar in the two channels.

As shown in section 5.7, apart from the small contribution from $B^\pm \to J/\psi\, \pi^\pm$ decays, the two main sources of background in the reference channel are the continuum (that smoothly crosses the signal region) and the partially reconstructed decays that dominate the low part of the mass spectrum, and that have a signal-like behaviour from the BDT point of view. Therefore, only the right part of the invariant mass distribution, almost exclusively dominated by the continuum background, can be used for this purpose.

To test the feasibility of this procedure the 2011 MC $b\bar{b} \to \mu^+\mu^- X$ is extremely useful since it allows to separate the different background components and therefore clearly understand their behaviour. Figures 6.6 and 6.7 report the comparison of the discriminating variable distributions for the 2011 MC $b\bar{b} \to \mu^+\mu^- X$ processed with the reconstruction algorithms for $B_s^0 \to \mu^+\mu^-$ and $B^\pm \to J/\psi\, K^\pm$ events in the mass region [5450,5630] MeV. The differences in the final state kinematic reflect on different distributions for ΔR, α_{2D}, decay length and p_T related variables that show how the $B^\pm$ right mass spectrum, given the current set of variables included, can not be used as a reliable sample for the MVA training and testing procedure. This option was

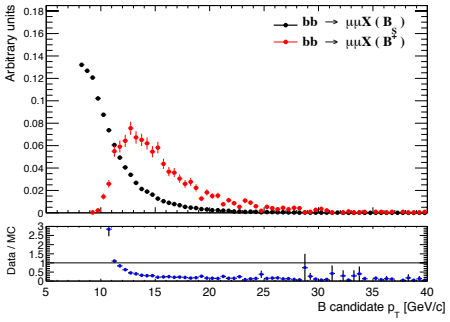


Fig. 6.6. *Comparison of the p_T of the B candidate for 2011 MC $b\bar{b} \to \mu^+\mu^- X$ reconstructed using the B_s^0 and $B^\pm$ algorithms. The black dots represent the B_s^0 reconstructed events while the red dots the $B^\pm$ reconstructed events.*

therefore not considered for the 2012 analysis and a specific MC (detailed in section 8.2.2) has been produced.

6.1.3 Additional Selection for $B_s^0 \to \mu^+\mu^-$ Background Reduction

The selection strategy for the $B_s^0 \to \mu^+\mu^-$ events adopted for the 2011 analysis (reported in chapter 5) is based on a two step approach: first of all a series of preliminary kinematic, track and vertex quality selections (summarised in section 5.4) provide a initial skim of the interesting events. In the second step, a set of discriminating variables, based on the topology of the event, are fed into a BDT algorithm to obtain a discriminating variable to be used for the final selection.

With a view of optimising the analysis and testing different invariant mass fit strategies (see section 7.2) another intermediate step has been added to the selection chain. The idea is to introduce a preliminary selection on the discriminating variables to reduce the combinatorial background maintaining high efficiency ($> 90\,\%$) for signal. This also allows to have a manageable number of events in the fits (useful to perform toy MC studies). From figure 5.7 it is clear how some variables show a sharp separation between the signal and the background regions and for three of them, pointing angle α_{2D}, ΔR and L_{xy} (reported in figure 6.8), the separation is extremely clean.

The selection (depicted also in figure 6.8) has been optimised in order to keep the signal efficiency well above $90\,\%$. Table 6.2 shows the results obtained applying the additional selection ($\alpha_{2D} < 1.0$, $\Delta R < 1.5$ and $\mathrm{L}_{xy} > 0$) that allow to maintain the signal efficiency at $95\,\%$, while reducing the background by roughly a factor 2.5.

Beside the rejection of the combinatorial background, the selection on α_{2D} and ΔR eliminates the peaks that these two variables have around π radians, and that are responsible for the double peak structure present in the left tail of the 2011 data BDT distribution (figure 5.22). A regular shape of the MVA output distribution can be easily parameterised using a simple function (like a Gaussian or an exponential), allowing to perform a 2-dimensional fit (mass vs BDT) for the signal extraction (see section 7.2.1).

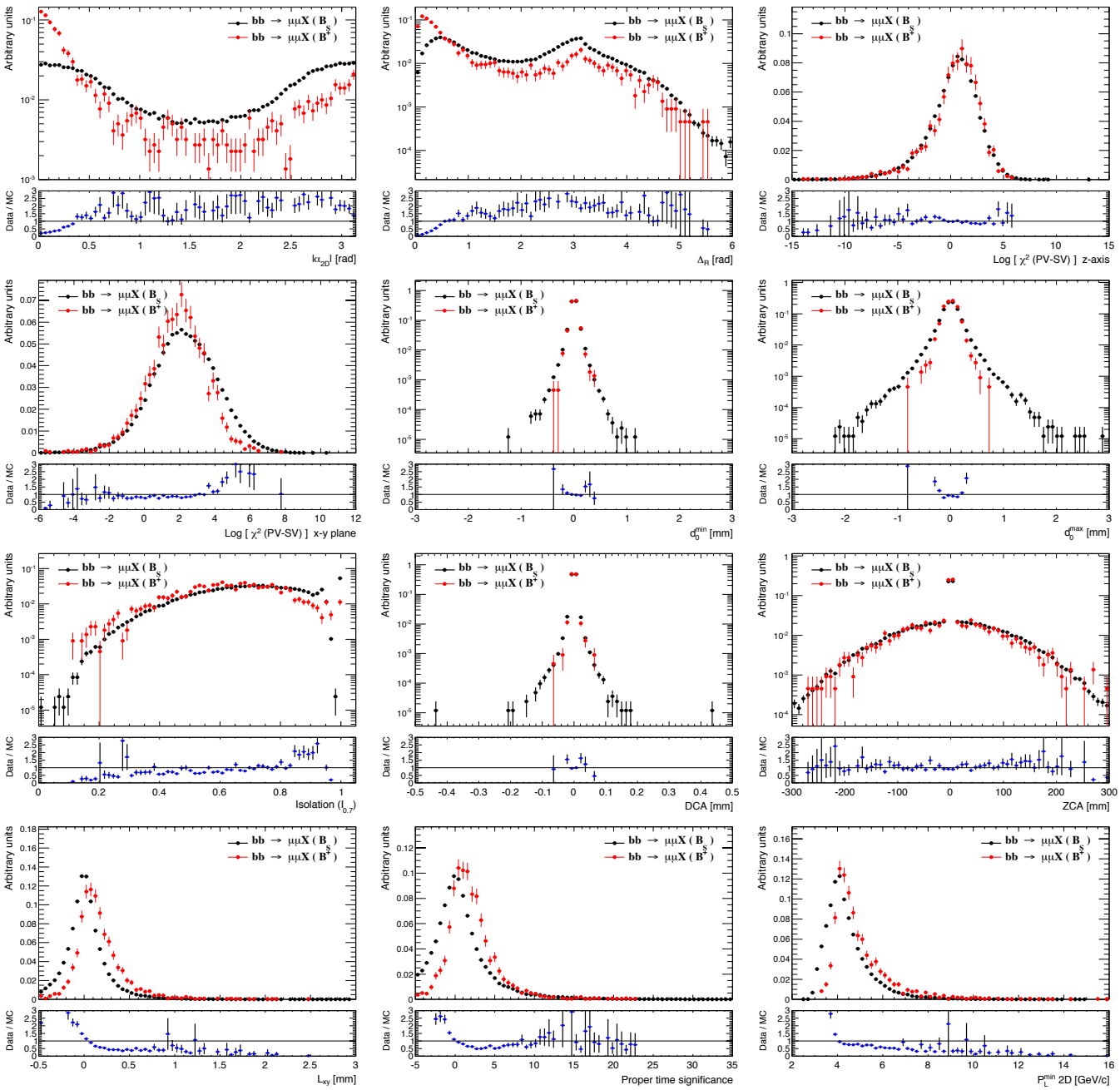


Fig. 6.7. *Comparison of the discriminating variables for 2011 MC $b\bar{b} \to \mu^+\mu^- X$ reconstructed using the B_s^0 and $B^\pm$ algorithms. The black dots represent the B_s^0 reconstructed events while the red dots the $B^\pm$ reconstructed events.*

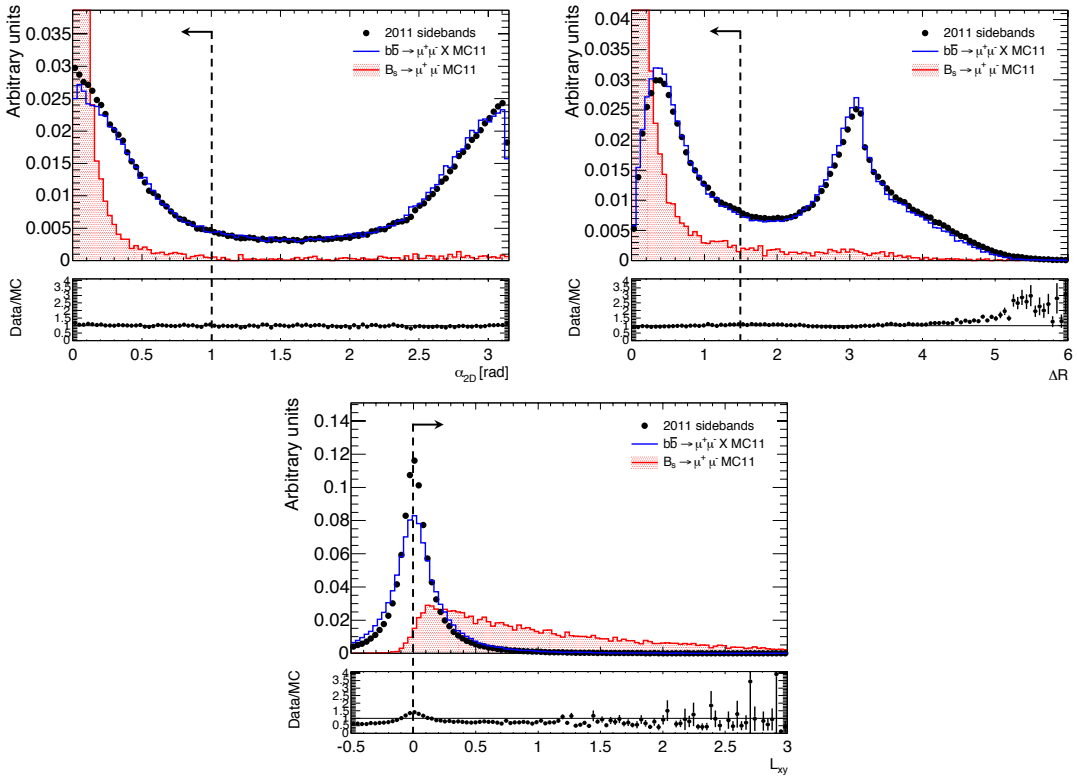


Fig. 6.8. *Distribution of the pointing angle α_{2D}, ΔR and L_{xy} for 2011 sideband data (black dots), 2011 MC $b\bar{b} \to \mu^+\mu^- X$ (blue solid line) and $B_s^0 \to \mu^+\mu^-$ signal (red-filled solid histogram). The black dashed vertical lines show the optimised selection.*

Selection	$B_s^0 \to \mu^+\mu^-$ efficiency	Background efficiency (sideband data)
$\alpha_{2D} < 1.0$	97 %	47 %
$\Delta R < 1.5$	96 %	44 %
$L_{xy} > 0$	97 %	53 %
Combined selection	95 %	42 %

Table 6.2. *Efficiency for 2011 MC $B_s^0 \to \mu^+\mu^-$ and 2011 sideband data obtained with the additional selection optimised to reduce the background. The combination of the selection on α_{2D}, ΔR and L_{xy} allows to maintain a high signal efficiency, reducing the background of a factor 2.5.*

6.2 Studies for the Signal Yield Extraction

Once the studies of the discriminating variables, the training of the multivariate analysis and the optimisation of the selection are completed, the final step is to extract

the number of $B_s^0 \to \mu^+\mu^-$ candidates from the observed mass distribution of events in data. In the past analysis this was obtained using a simple cut and count method. In this section an approach based on the fit of the di-muon mass distribution, that allows to extract the number of signal and background events in data, is reviewed.

To estimate the value of a physics quantity in data we need to formulate a model that describes the expected distributions of a series of observables $\boldsymbol{x}$ given a set of parameters of interest $\boldsymbol{p}$. This process requires a definition of a test statistic T that, given an observed data sample $\boldsymbol{x_0}$, can be used to infer a statement about the parameter $\boldsymbol{p}$ for which the observed data distribution is most likely. The most common test statistic used in particle physics are the χ^2 and the likelihood. In both cases, the fit in data corresponds to an optimisation process in the parameter space $\boldsymbol{p}$ that maximises (minimises) the likelihood (χ^2) values. In the next sections we will focus on the likelihood estimation.

6.2.1 Probability Density Functions

The models for the distributions of the various observables can be described using a positive-defined probability density function (p.d.f.)

$$\mathcal{M}(\boldsymbol{x}; \boldsymbol{p}, \boldsymbol{q}) \quad , \tag{6.1}$$

where $\boldsymbol{x}$ are the observable quantities, $\boldsymbol{p}$ the parameter(s) of interest and $\boldsymbol{q}$ is a set of additional "nuisance parameters" that represent quantities, such as the mass resolution (see section 8.4), that affect the relation between $\boldsymbol{p}$ and $\boldsymbol{x}$ and have to be inferred from data, not being known a priori. The probability density function is normalised to unity over the range of the observable $\boldsymbol{x}$ and for any value of $\boldsymbol{p}$ and $\boldsymbol{q}$:

$$\forall\, \boldsymbol{p}, \boldsymbol{q} \;:\; \int \mathcal{M}(\boldsymbol{x}; \boldsymbol{p}, \boldsymbol{q})\, d\boldsymbol{x} = 1 \quad . \tag{6.2}$$

6.2.2 Maximum Likelihood Estimation

The basis of the parameter inference using a model $\mathcal{M}$ and observed data is the *likelihood*, that is defined as the probability density function evaluated for the value of the variable $\boldsymbol{x}$ for a specific event i,

$$\mathcal{L}_{\boldsymbol{x_i}}(\boldsymbol{p}, \boldsymbol{q}) = \mathcal{M}(\boldsymbol{x_i}; \boldsymbol{p}, \boldsymbol{q}) \quad . \tag{6.3}$$

For example, considering our model made of a background and a signal component, if the event $\boldsymbol{x_i}$ is certainly signal, then the part of $\mathcal{M}(\boldsymbol{x_i})$ corresponding to signal will be one, similarly if the event is definitely not signal then the part of $\mathcal{M}(\boldsymbol{x_i})$ corresponding to signal will be zero. Given a measurement, made of N points (events), the total un-binned likelihood is simply the product of the likelihood of each observation

$$\mathcal{L}(\boldsymbol{p}, \boldsymbol{q}) = \prod_{i=1}^{N} \mathcal{M}(\boldsymbol{x_i}; \boldsymbol{p}, \boldsymbol{q}) \quad . \tag{6.4}$$

In order to simplify the numerical evaluation, when the amount of data is relevant, often the Negative Log-Likelihood (NLL),

$$-\log \mathcal{L}(\boldsymbol{p}, \boldsymbol{q}) = -\sum_{i=1}^{N} \log \mathcal{M}(\boldsymbol{x_i}; \boldsymbol{p}, \boldsymbol{q}) \quad , \tag{6.5}$$

is adopted[1] instead of the likelihood defined in 6.4. The value of the parameter $\boldsymbol{p}$ that maximises the likelihood (or equivalently minimises the NLL) is the parameter estimator $\hat{\boldsymbol{p}}$.

When the number of measurements is very large, in order to reduce the processing time, it is possible to minimise a binned NLL

$$-\log \mathcal{L}(\boldsymbol{p}, \boldsymbol{q}) = -\sum_{i=1}^{N} n_i \cdot \log \mathcal{M}(\boldsymbol{x_i}; \boldsymbol{p}, \boldsymbol{q}) \quad , \tag{6.6}$$

where $\boldsymbol{x_i}$ and n_i represent respectively the bin centre and the number of measurements contained in the bin i for a histogram with N bins. If the width of the bins is small, then the loss of precision is negligible.

The properties of likelihood estimators are extensively described in literature [130]. The maximum likelihood criterion is an unbiased and efficient estimator if the statistics is sufficiently large, whereas particular attention should be paid in case of very small (signal) event counts N_{obs} where biased terms, generally proportional to $1/N_{obs}$, can appear in the likelihood and may be non-negligible compared to the statistical uncertainty, approximately proportional to $1/\sqrt{N_{obs}}$.

6.2.3 The Extended Maximum Likelihood Formalism

In an analysis we can use p.d.f. models to describe samples in terms of different categories. The simplest model, that allows to extract the signal yield from data in the presence of background, is constituted by a sum of signal and background components:

$$\mathcal{M}(\boldsymbol{x_i}; \boldsymbol{p}, \boldsymbol{q}) = \beta \cdot \mathcal{S}(\boldsymbol{x_i}; \boldsymbol{p}) + (1 - \beta) \cdot \mathcal{B}(\boldsymbol{x_i}; \boldsymbol{q}) \quad , \tag{6.7}$$

where $\mathcal{S}(\boldsymbol{x_i}; \boldsymbol{p})$ and $\mathcal{B}(\boldsymbol{x_i}; \boldsymbol{q})$ are respectively the signal and the background distributions of the observable $\boldsymbol{x_i}$, and β is the fraction of signal in the data. In this way, the parameters of $\mathcal{S}(\boldsymbol{x_i}; \boldsymbol{p})$ and $\mathcal{B}(\boldsymbol{x_i}; \boldsymbol{q})$ can be constrained from data, and the number of signal events is β times the total number of events. Since in the model in equation 6.7 the p.d.f. function can only describe the shape of the distribution of the observed events and not its count, we need a different method that allows to determine the event yields of a set of categories (in our case signal and background) instead of a fraction of events.

[1] Applying the natural log function in this context is useful for several reasons. First, numerical analysis reminds us that logs reduce potential for underflow, due to very small likelihoods. Second, calculus reminds us logs permit the additional trick: converting a product of factors into a summation of factors. Finally, calculus again reminds us that the natural log function is a monotone transformation, thus the extrema of $\mathcal{L}$ are equivalent to the extrema of $\log \mathcal{L}$.

In the Extended Maximum Likelihood formalism (EML) [131, 132] the normalisation of the model is not fixed to one, but to the parameter $N_{\rm exp}$ and the likelihood given by 6.4 becomes

$$\mathcal{L}(\boldsymbol{p},\boldsymbol{q}) = \text{Poisson}\left(N_{\rm obs} \,|\, N_{\rm exp}(\boldsymbol{p},\boldsymbol{q})\right) \cdot \prod_{i=1}^{N} \mathcal{M}(\boldsymbol{x_i};\boldsymbol{p},\boldsymbol{q}) \quad . \tag{6.8}$$

where N_{obs} is the observed event count modelled by a Poisson distribution with mean equal to the expected event count $N_{\rm exp}(\boldsymbol{p},\boldsymbol{q})$. The likelihood of the composite model can then be rewritten in the EML formalism taking

$$\mathcal{M}(\boldsymbol{x_i};\boldsymbol{p},\boldsymbol{q}) = \frac{N_s}{N_s+N_b} \cdot \mathcal{S}(\boldsymbol{x_i};\boldsymbol{p}) + \frac{N_b}{N_s+N_b} \cdot \mathcal{B}(\boldsymbol{x_i};\boldsymbol{q}) \tag{6.9}$$

as the probability density function and

$$N_{\rm exp} = N_s + N_b \tag{6.10}$$

as the number of expected event count (for our purposes, the observable quantity $\boldsymbol{x}$ is the di-muon invariant mass). A minimisation of the EML will return the estimates for the signal (N_s) and background (N_b) event yields.

6.2.4 Mass Fit Tests

The performance and the reliability of the invariant mass fit procedure have been tested by means of mock data (toy MC experiments), generated using the information obtained from 2011 sideband and MC signal events. In section 6.3 it will also be shown how the invariant mass fit can be reliably applied to the un-blinded 2011 dataset, providing results that are compatible with the public analysis [106].

The code built for all the tests reported in the next sections is based on the `Roofit` toolkit (Werkerke and Kirkby, 2003 [134]), developed by the *BABAR* Collaboration, that allows to build p.d.f. of arbitrary complexity inside the ROOT framework with a minimum amount of code. The likelihood is minimised using the standard Minuit heuristic algorithm `MIGRAD` that searches for minima mostly following a strategy based on a steepest descent algorithm, numerically calculating the gradient of the input functions.

Using the extended maximum likelihood formalism, the number of signal and background events in the mass region [4766,5966] MeV has been extracted from 2011 data after having applied the optimised BDT selection for the 2011 analysis. The signal distribution has been modelled using a double Gaussian

$$\mathcal{S}(\boldsymbol{m};\mu,\sigma_1,\sigma_2,k) = k \cdot e^{-\frac{(\boldsymbol{m}-\mu)^2}{2\sigma_1}} + (1-k) \cdot e^{-\frac{(\boldsymbol{m}-\mu)^2}{2\sigma_2}} \quad , \tag{6.11}$$

where k is a factor that expresses the relative contribution of the two Gaussians and $\boldsymbol{m}$ is the di-muon invariant mass of the events. A single Gaussian is not able to accurately describe the shape since the mass resolution depends on the region of the detector

crossed by the muons (see section 8.4 for more details). The sideband data, used to describe the continuum background expected in the signal region, can be fitted using an exponential function

$$\mathcal{B}(\boldsymbol{m};\alpha) = e^{\alpha\cdot\boldsymbol{m}} \quad . \tag{6.12}$$

The available statistic in the di-muon mass region used for the fit allows to extract the exponential shape directly from data and thus the α parameter is not fixed in fit. The same is not valid for the signal, since only very few events are expected in data, and the shape parameters (μ, σ_1, σ_2 and k) are therefore fixed to the MC values.

The number of background and signal events can be finally estimated minimising the negative log-EML

$$-\log\mathcal{L}(\boldsymbol{m};\alpha) = -\text{Poisson}\left(N_{\text{obs}} \,|\, N_{\text{exp}}(\alpha)\right) - \log\sum_{\text{data}}\mathcal{M}(\boldsymbol{m};\alpha) \quad , \tag{6.13}$$

where the p.d.f. model is described by 6.9.

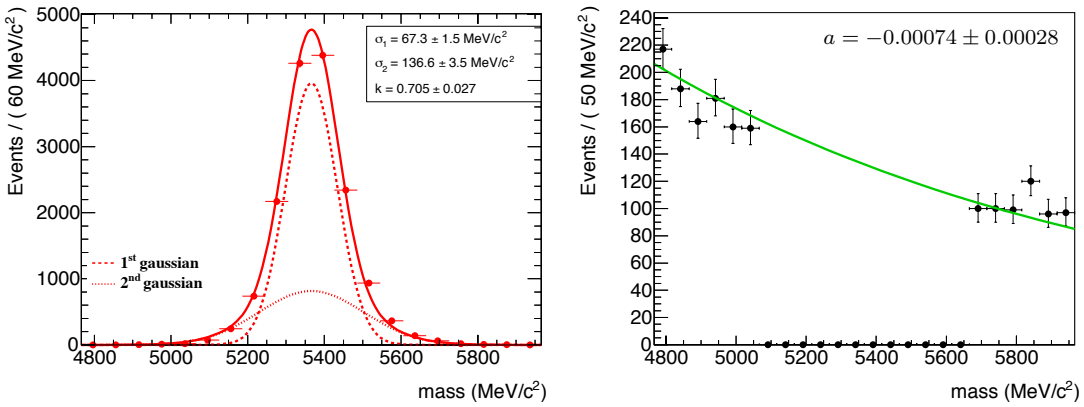


Fig. 6.9. *Left: fit performed on 2011 MC signal using a double Gaussian function and without applying the BDT selection. Right: example of fit on 2011 blinded data using an exponential function and requiring BDT > 0.05.*

As already stated, toy MC experiments are generated using the shape parameters obtained from the fit of 2011 MC signal and sideband data[2]. Figure 6.9 shows the mass fit performed on 2011 $B_s^0 \to \mu^+\mu^-$ signal MC without applying any BDT selection and an example of exponential fit on sideband data applying an arbitrary BDT selection.

6.2.4.1 Early Tests on 2011 Datasets

The first series of studies reported in this section has been done generating the number of events estimated using a preliminary version of the BDT and a very preliminary event selection available at that time. The total number of signal events expected was ~ 4. Although an optimised MVA output is undoubtedly mandatory

[2]See section 5.3 for more information about the datasets used.

to improve the signal sensitivity, for these very preliminary tests and the purpose of studying the intrinsic fit performance it is not fundamental.

The shape of the signal model (μ, σ_1, σ_2, and k) is fixed to the MC shown in figure 6.9 while the exponential parameter α, the number of signal (N_s) and background (N_b) events are free to float in the fit. Figure 6.10 shows an example of un-binned EML fit, using the p.d.f. described above, on one single toy experiment where the number of signal and background events generated are respectively 1 and 72. The signal yield obtained in this example is 4.9 ± 5.1 whereas the background yield is 68.1 ± 9.4.

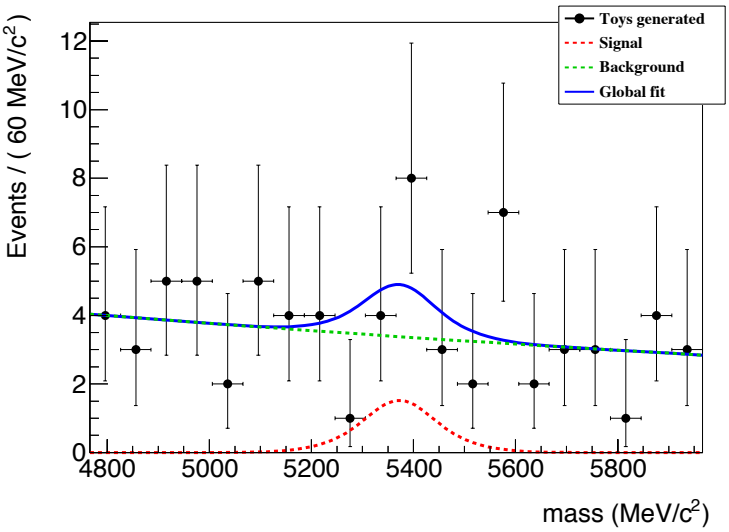


Fig. 6.10. *Un-binned EML fit performed on a single toy experiment where the number of signal and background events generated are respectively 1 and 72. The black dots are the data generated, the dashed red and green lines are respectively the signal and background components, while the blue line is the global fit.*

Of course, one single test is not statistically meaningful. Maximum likelihood fits on complex models can be validated by studying their behaviour simulating many ($\mathcal{O}(1000)$) data samples according to the model considered (fluctuating the number of signal and background events accordingly to a Poisson distribution), and fitting the model to each of these datasets. For each generated sample, the pull of the estimated parameters, defined as

$$\mathrm{PULL}(p) = \frac{\hat{p}_{\,\mathrm{fit}} - p_{\,\mathrm{true}}}{\sigma_{\hat{p}\,\mathrm{fit}}} \tag{6.14}$$

is evaluated. If the mean of the pull distribution obtained is consistent with zero, the estimator $\hat{p}$ is free of bias, i.e. it estimates the true value correctly on average. If the variance of the pull distribution is consistent with one, the estimator $\sigma_{\hat{p}}$ correctly represents the uncertainty. Therefore, a too narrow pull distribution is sign that $\sigma_{\hat{p}}$ overestimates the errors, while a too wide distribution indicates an underestimated uncertainty. Understanding and quantifying the potential bias is quite important for estimators of small yields for which the EML can give a significant biasing effect. It is

important to emphasise that, when we generate an ensemble of toy MC experiments, we need to vary the number of generated events of each species according to a Poisson distribution with a mean equal to the corresponding expected number of events, otherwise the width of the fitted distribution will not be statistically meaningful.

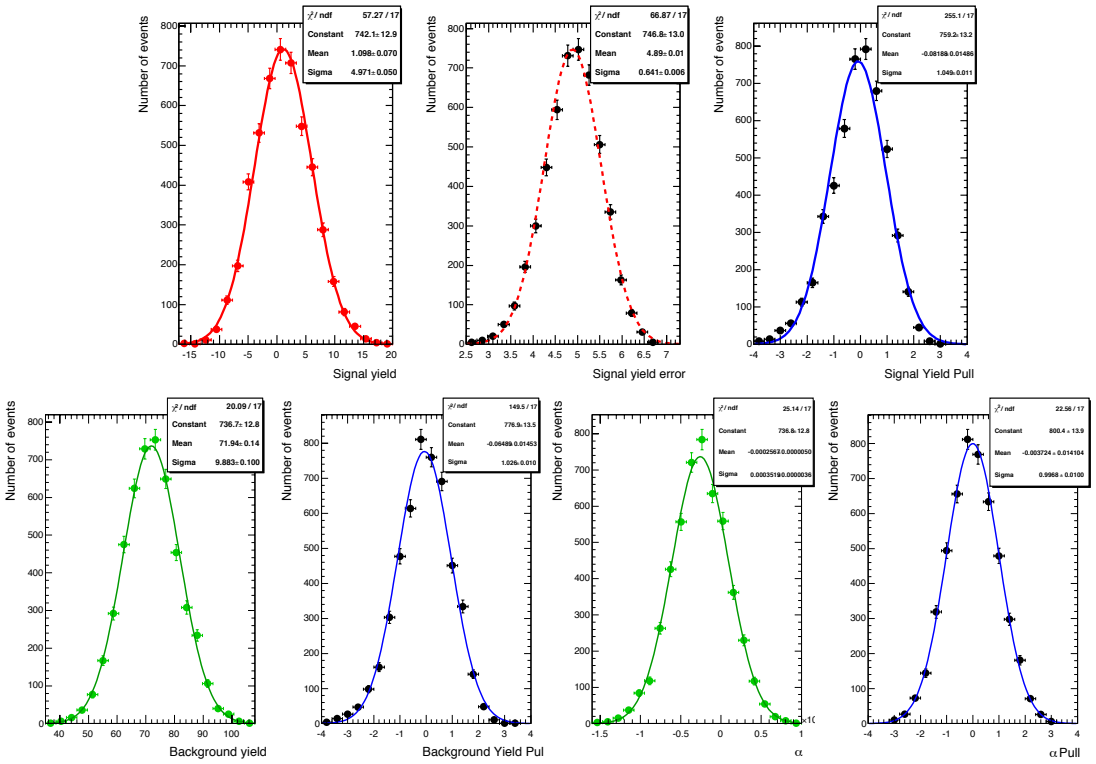


Fig. 6.11. *Signal yield (top), background yield (bottom left) and α parameter (bottom right) distributions, with the relative pulls, obtained from 5000 toy MC experiments. The number of signal and background events generated are respectively 1 and 72.*

The fit performance have been tested analysing the pull distributions for signal and background yields as well as for the α parameter of the exponential function. Toy MC experiments have been generated and fitted with different numbers of background and signal events, obtained varying the BDT selection cut. The number of signal events expected for a certain BDT selection has been obtained rescaling the total expected signal events by the efficiency associated to the BDT selection

$$N_s = 4 \cdot \frac{N_s^{MC}(\text{BDT} > \text{BDT}_{cut})}{N_s^{MC}(\text{NO BDT selection})} \tag{6.15}$$

and then rounding the obtained value to the closest integer. Table 6.3 summarises the results obtained generating 5000 toys MC experiments in each configuration, whereas figure 6.11 shows an example of the yield and pull distributions of the fitted variables obtained from 5000 toy MC experiments generating an average of 1 signal and 72

$\mathbf{BDT}_{cut}$	**0.08**	**0.1**	**0.12**
N_s generated	2	2	2
N_b generated	1158	528	278
N_s	1.3 ± 15.1	1.7 ± 12.2	1.6 ± 9.5
N_s pull mean	-0.028 ± 0.014	-0.062 ± 0.013	-0.033 ± 0.015
N_s pull variance	0.98 ± 0.01	0.99 ± 0.1	1.00 ± 0.01
N_b	1158 ± 38	529 ± 25	278 ± 19
N_b pull mean	-0.029 ± 0.012	-0.011 ± 0.012	-0.022 ± 0.014
N_b pull variance	0.98 ± 0.01	0.96 ± 0.01	0.98 ± 0.01
α	-0.0008 ± 0.0001	-0.0009 ± 0.0001	-0.0009 ± 0.0002
α pull mean	0.019 ± 0.014	0.011 ± 0.015	-0.003 ± 0.014
α pull variance	0.99 ± 0.1	1.00 ± 0.01	0.99 ± 0.01

$\mathbf{BDT}_{cut}$	**0.14**	**0.16**	**0.18**
N_s generated	1	1	1
N_b generated	175	107	72
N_s	1.1 ± 8.0	1.3 ± 6.7	0.6 ± 5.1
N_s pull mean	-0.088 ± 0.020	-0.037 ± 0.024	-0.070 ± 0.016
N_s pull variance	1.00 ± 0.01	0.98 ± 0.01	0.98 ± 0.01
N_b	175 ± 15	107 ± 12	72 ± 10
N_b pull mean	-0.031 ± 0.009	-0.055 ± 0.020	-0.061 ± 0.022
N_b pull variance	0.97 ± 0.01	0.99 ± 0.01	0.98 ± 0.01
α	-0.0007 ± 0.0002	-0.0005 ± 0.0003	-0.0003 ± 0.0004
α pull mean	0.004 ± 0.014	0.03 ± 0.01	0.004 ± 0.014
α pull variance	0.99 ± 0.01	0.99 ± 0.01	1.00 ± 0.01

Table 6.3. *Results obtained generating 5000 toy MC experiments for various BDT selections.*

background events. The variance of the pull distributions is always compatible with one, sign that the errors are treated correctly. On the contrary, the mean values are not perfectly compatible with zero, which translates into a bias in the values of the yields obtained. This effect is related to the asymmetric tails of the distributions (see figure 6.11) and it can be explained with the very low statistic of the expected signal yield. To prove that, toy MC experiments have been generated considering an expected number of signal events equal to zero, or artificially increasing both signal and background available statistics by a factor 20. The results are summarised in table 6.4 and show, in agreement with what expected, that the maximum bias (defined as the product of the pull times the variance of the signal yield, $\text{PULL} \cdot \sigma_{N_s^{fit}}$) is measured when the number of expected signal events is null and that, as the number of events in

the sample increases, the relative fit bias on the yield scales down.

These tests illustrate that the fit procedure is correct and reliable, and the bias on the signal yield is related to small-numbers fluctuations.

	$\mathbf{BDT}_{cut}$	N_s^{GEN}	N_s^{fit}	N_s^{fit} PULL	PULL $\cdot\, \sigma_{N_s^{fit}}$
4.5 fb^{-1}	0.08	2	1.3 ± 15.1	-0.028 ± 0.014	-0.42 [21.0 %]
	0.12	2	1.6 ± 9.5	-0.033 ± 0.015	-0.31 [15.5 %]
	0.18	1	0.6 ± 5.1	-0.070 ± 0.016	-0.36 [36.0 %]
4.5 fb^{-1} $N_s = 0$	0.08	0	(-0.005) ± 16.5	-0.029 ± 0.014	-0.48
	0.12	0	(-0.09) ± 9.1	-0.048 ± 0.015	0.44
	0.18	0	(0.1) ± 5.3	-0.075 ± 0.016	0.40
90 fb^{-1}	0.08	35	32.6 ± 84.9	-0.032 ± 0.015	-2.7 [7.7 %]
	0.12	23	22.1 ± 42.1	-0.025 ± 0.014	-1.1 [4.8 %]
	0.18	11	10.5 ± 21.6	-0.032 ± 0.014	-0.7 [6.4 %]

Table 6.4. *Results obtained generating 5000 toy MC experiments simulating a statistic of 4.5 fb^{-1}, 4.5 fb^{-1} with $N_s = 0$, and 90 fb^{-1}. For each test the signal yield, the related pull and the signal bias are shown.*

6.2.5 Mass Fit Optimisation on MC Background

After having tested the reliability of the invariant mass fit, the following step is to optimise the strategy for obtaining the maximum significance. The procedure explained in section 5.6 is optimised for a "cut & count" analysis, but as we intend to investigate the possibility of performing an un-binned maximum likelihood fit, a different strategy has to be investigated. The results reported in this section and in the next ones have been obtained using the final version of data and MC samples as well as of the BDT classifier used for the public 2011 analysis.

The BDT selection used in the 2011 analysis, removes most of the continuum background, but also significantly reduces the number of signal events available. Using the mass fit, that takes also advantage from the knowledge of the shape of the signal and background components, the idea is to try to relax the BDT selection in order to increase the efficiency on the signal.

As shown in the previous section, the estimation of the fit performance is linked to the study of the behaviour of quantities like the pull and the bias on the event yields. These quantities are strictly related to the absolute number of signal and background events, that in turn are related to the chosen BDT selection. The new strategy explained in this section is based on a scan of the BDT output, trying to find an optimal working point in terms of the above quantities. In order to avoid biases that could rise from using the same sample for the optimisation of the fit and the extraction of the signal yield, the procedure is entirely done on the 2011 MC $b\bar{b} \to \mu^+\mu^- X$. The signal and background shapes have been modelled using respectively a double

Gaussian and an exponential (equation 6.11 and 6.12) as fitted from 2011 MC samples, the expected number of signal events has been estimated using the formula in 6.15, while for the background a parameterisation of the right tail of the BDT distribution has been introduced. From figure 6.12 it is clear how, for higher values of the BDT output, the estimation of the number of expected background events is dominated by statistical fluctuations. In order to obtain a smooth shape, the right tail of the 2011 MC background $b\bar{b} \to \mu^+\mu^- X$ has been parameterised as a function of the BDT output, using a combination of an exponential and a hyperbola

$$\mathcal{T}(\mathrm{BDT}; l, m, n) = l \cdot e^{-m \cdot \mathrm{BDT}} + \frac{n}{\mathrm{BDT}} \quad . \tag{6.16}$$

The background events selected by a certain BDT value are extracted from the shape obtained from the fit and rescaled to data considering the different data-MC normalisation.

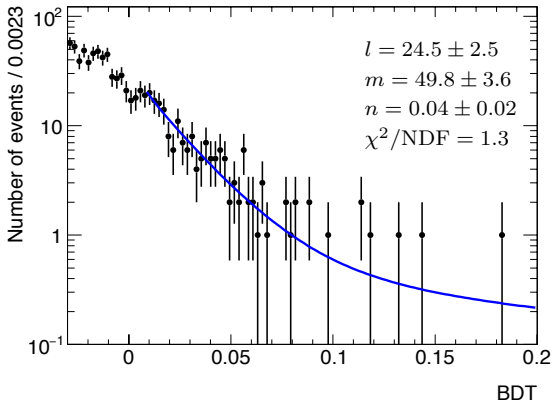
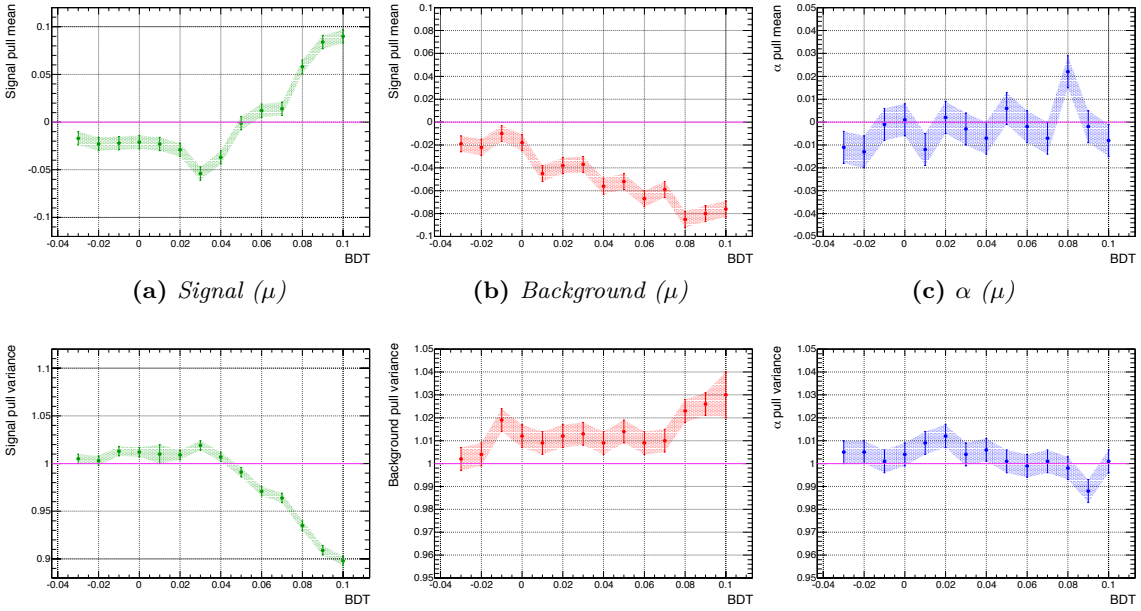


Fig. 6.12. *Right tail of the BDT distribution of the 2011 MC $b\bar{b} \to \mu^+\mu^- X$ sample. The solid blue line is the fit obtained considering a combination of an exponential and a hyperbola function (parameters described in the text).*

The scan has been performed in the BDT output interval [-0.03 , 0.1] with steps of 0.01. Three quantities have been used to assess the performance of the fit: the signal error, the signal pull and the signal bias. Figure 6.13 shows the behaviour of the pull for the signal yield, the background yield and the background slope, while table 6.5 reports their values obtained from the scan. As shown in figure 6.14, the optimal BDT value of 0.045 corresponds to the working point that simultaneously minimises the three quantities. This selection has a signal efficiency of 58 % (the efficiency of the BDT selection optimised for the 2011 analysis was $\sim 25\,\%$).

Since the agreement of data and MC BDT distributions is not optimal, the result obtained can not be reliably applied to data. The optimised selection on MC events has been reported into the corresponding one on data by means of the well known *signal-to-noise* estimator. Once the value of this estimator for the obtained selection has been evaluated on the MC, a scan on data has been performed looking for the



(a) *Signal (μ)* **(b)** *Background (μ)* **(c)** *α (μ)*

(d) *Signal (σ)* **(e)** *Background (σ)* **(f)** *α (σ)*

Fig. 6.13. *Mean and variance of the pull distributions for the signal yield (left), background yield (middle) and background slope α (right) obtained considering the 2011 statistic.*

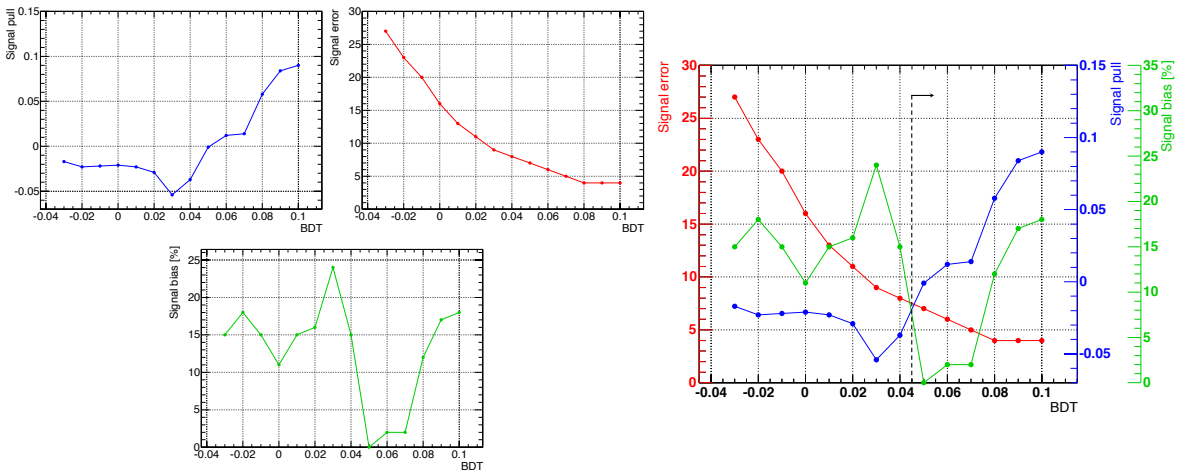


Fig. 6.14. *Behaviour of the most significant fit parameters for a BDT scan in the range [-0.03,0.1] (left) and combination of the three results (right), considering the 2011 statistic. The optimal working point, underlined by the black dashed vertical line, corresponds to a cut of 0.045.*

BDT selection that gives exactly the same value of *signal-to-noise*. The optimised BDT selection obtained on data has been found to be equal to 0.08, corresponding to an estimated signal efficiency of 45 %.

$\mathbf{BDT}_{cut}$	N_s^{GEN}	N_s^{err}	N_s pull mean	N_s bias [%]
-0.03	3	27	-0.017	15
-0.02	3	23	-0.023	18
-0.01	3	20	-0.022	15
0	2	16	-0.021	11
0.01	2	13	-0.023	15
0.02	2	11	-0.029	16
0.03	2	9	-0.054	24
0.04	2	8	-0.037	15
0.05	2	7	-0.001	0
0.06	2	6	0.012	2
0.07	2	5	0.014	2
0.08	1	4	0.058	12
0.09	1	4	0.084	17
0.10	1	4	0.090	18

Table 6.5. *Results of the BDT scan considering the 2011 statistic.*

The selection has been finally tested on 2011 un-blinded data giving a result of 1.3 ± 5.8 signal events (~ 2 events are expected) with 119 ± 12 background events. The obtained fit is reported in figure 6.15.

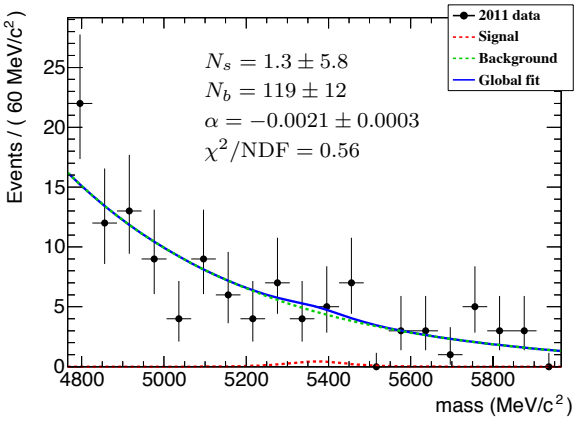
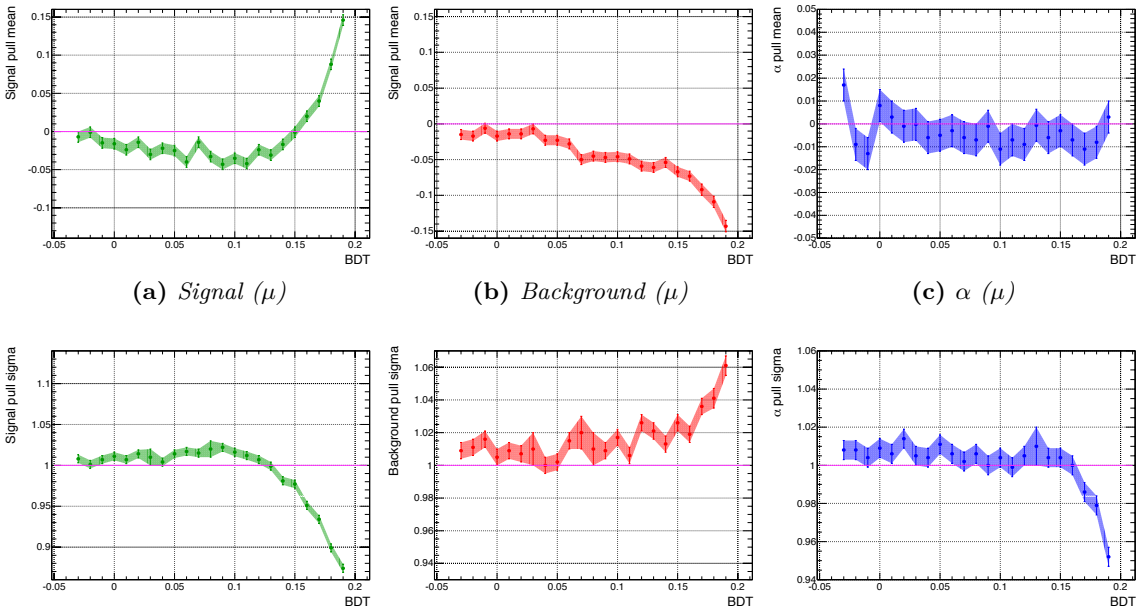


Fig. 6.15. *Invariant mass fit on the 2011 un-blinded data for BDT > 0.08 optimised on 2011 MC events (the parameter α is the slope of the exponential function).*

The same procedure has been tested generating the statistic expected in 2012. At the c.m. energy of 8 TeV, the $B_s^0 \to \mu^+\mu^-$ cross-section is expected to increase up to 15 % and the number of signal events (with no BDT selection) is estimated to be around 18 for an integrated luminosity of 20 fb^{-1} (in this estimation we are considering 4 events expected in 2012 without the BDT selection[3]).

[3]The number of expected events in 2012 is obtained multiplying the number of events expected in



(a) *Signal (μ)* **(b)** *Background (μ)* **(c)** *α (μ)*

(d) *Signal (σ)* **(e)** *Background (σ)* **(f)** *α (σ)*

Fig. 6.16. *Mean and variance of the pull distributions for the signal yield (left), background yield (middle) and background slope α (right) simulating a statistic of 20 fb^{-1}.*

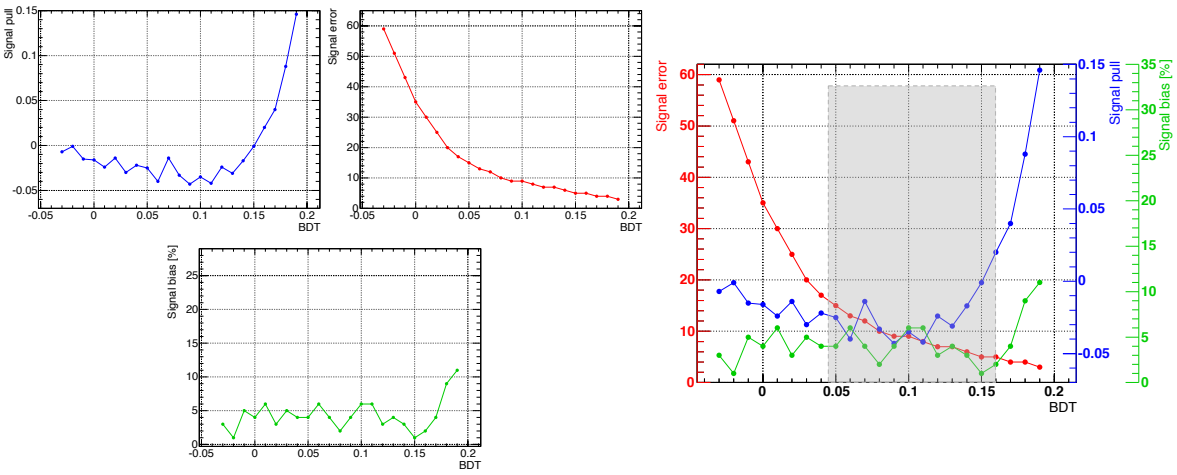


Fig. 6.17. *Behaviour of the most significant fit parameters for a BDT scan in the interval [-0.03,0.2] (left) and combination of the three results (right) simulating a statistic of 20 fb^{-1}. The shadowed grey rectangle, [0.045-0.16], shows the optimal region in which the fit gives the best performance.*

The mean and variance of the pull distributions for the signal and background

2011 (4) by the luminosity factor (20 fb^{-1}/ 5 fb^{-1} = 4) and increasing the result by 15 % in order to take into account of the enhanced cross-section.

yields, and the background shape α are reported in figure 6.16, while figure 6.17 shows the parameters behaviour for the different values of the scan. In this case, instead of a single working point, it is possible to identify a region, [0.045-0.16], in which the mass fit gives the best performance. The increased statistic allows to reduce the bias on the signal yield, leading to a pull distribution that is very compatible with zero for a wide range of BDT values. The importance of this will be more clear when the new strategies for the signal yield extraction will be treated (see section 7.2).

6.2.6 Combinatorial Background: Exponential Versus Linear Function

In the previous sections the default modelling of the combinatorial background was the exponential function. The reason is mostly "historical" rather than related to the physics of the processes under study. From the plots shown in the previous sections, as well as in section 6.1.1, we can see how the combinatorial background can be modelled using also a linear function. This is confirmed in figure 6.18 where the un-blinded 2011 data (without any BDT selection applied) have been fitted using an exponential and a linear model. The identical χ^2/NDF confirms how both functions can be reliably used to extract the shape from data.

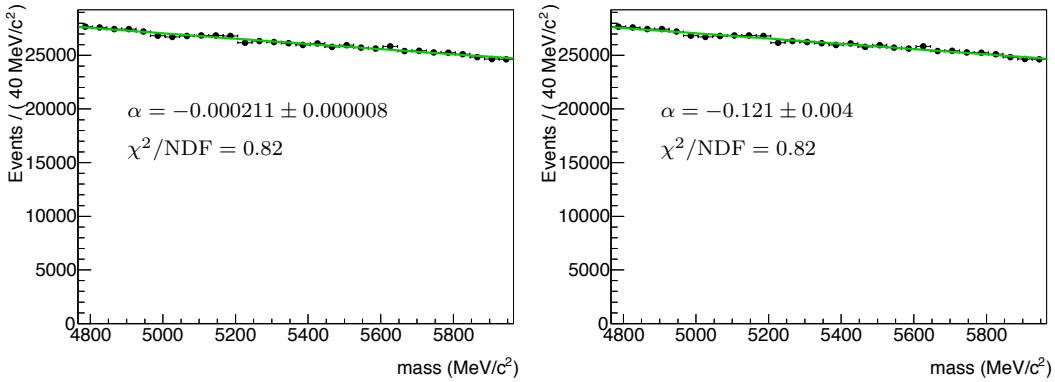


Fig. 6.18. *Comparison of the invariant mass fit performed using an exponential (left) and a linear (right) function on 2011 un-blinded data without applying any BDT selection (the parameter α is the slope of the two fitting functions).*

Detailed studies on the background mass dependence will be shown in chapter 8, but as anticipated in section 6.1.1, the different behaviour of the background for higher values of the BDT output (where the combinatorial background is highly rejected) is related to different (signal-like) background components that populate the left mass sideband and that can be modelled using an exponential. Although this is not so obvious in 2011 data due to the limited statistic available that do not allow to properly and reliably model the shape of these contributions, the 2012 data will help to better understand the composition of this background that will be extensively analysed in sections 8.10.2 and 8.10.3.

In the previous sections, it has been shown how the fit procedure is stable and reliable, and how it is possible to optimise the BDT selection directly on MC, avoiding introducing any kind of bias in the signal yield extraction. The last step in the fit validation procedure is to evaluate the upper limit on 4.9 fb^{-1} of data considering both the BDT selection of the 2011 analysis as well as the one obtained from the above optimisation. The next section will review the strategy to evaluate the upper limit using the CL_s method with the un-binned maximum likelihood fit. The result obtained on 2011 un-blinded dataset with the mass fit will be compared to the one extracted with the cut and count analysis. In order to try also to quantify the dependence of the final results on the background modelling, these tests will be performed shaping the combinatorial background using both an exponential and a linear function.

6.3 CL_s Method for the Upper Limit Extraction

It is quite common in particle physics to search for processes that have been predicted, but not seen yet. For the purpose of discovering a new signal process, one defines the null hypothesis (H_0) that describes only known processes (background), and that is tested against the alternative hypothesis (H_1) that includes both the background and the expected signal. When setting limits, the model with *signal+background* plays the role of H_0, which is tested against the *background-only* hypothesis, H_1.

When we analyse our data, one of the first step is to evaluate the compatibility of the observed data with the H hypothesis under study. This is quantified by the p-value of the data which gives the probability, under the assumption of H, of finding data of equal or worse agreement with the prediction of H. Large p-values therefore correspond to datasets that agree well with the H hypothesis, while one can regard H as excluded if the p-value is below a specified threshold. In particle physics the p-value is usually converted into the equivalent *significance*

$$Z = \Phi^{-1}(1-p) \quad , \tag{6.17}$$

defined such that a Gaussian distributed variable, found Z standard deviations above its mean, has an upper-tail probability equal to p, where Φ^{-1} is the quantile (inverse of the cumulative distribution) of the standard Gaussian. To claim for a discovery, the particle physics community usually considers the rejection of the background hypothesis with a significance of at least $Z = 5$ that corresponds to a p-value of $2.87 \cdot 10^{-7}$, while for the purpose of excluding a signal hypothesis, a threshold p-value of 0.05 (i.e. 95 % confidence level) is quite often used and corresponds to $Z = 1.64$. The *sensitivity* of an experiment can be quantified by reporting the expected (median or mean) significance that one would obtain for a variety of signal hypotheses.

In this section we will analyse a widely used procedure to establish discovery (or exclusion) in particle physics, based on a frequentist significance test using a likelihood ratio as a test statistic. Defining the *signal strength* parameter

$$\mu = \frac{N_{signal}}{N^{SM}_{signal}} \quad , \tag{6.18}$$

such that $\mu = 0$ is the *background-only* model while $\mu = 1$ corresponds to the nominal SM signal model, we can evaluate, for a certain hypothesis μ, the profile likelihood ratio

$$\lambda(\mu) = \frac{\mathcal{L}(\mu, \hat{\hat{\theta}}(\mu))}{\mathcal{L}(\hat{\mu}, \hat{\theta})} \quad , \tag{6.19}$$

where θ represents parameters that characterise the shapes of the p.d.f.s. $\lambda(\mu)$ allows to quantify the level of agreement between hypothesis and data. The $\hat{\hat{\theta}}$ in the numerator denotes the value of θ that maximises the likelihood for a specified μ, while at denominator the likelihood is maximised with respect to both μ and θ, thus $\hat{\mu}$ and $\hat{\theta}$ are the maximum likelihood estimators. If the tested μ is close to the value $\hat{\mu}$ obtained by the fit on data and so also the parameters $\hat{\theta}$ are close to $\hat{\hat{\theta}}$ that maximises the numerator, the profile likelihood ratio is equal to the unity. On the other hand, if the hypothesis is not compatible with the observed data, the ratio tends to zero.

From the definition of $\lambda(\mu)$ we can create three different tests. The first one is simply defined as

$$q_\mu^{\text{basic}} = -2 \cdot \log \frac{\mathcal{L}(\mu, \hat{\hat{\theta}}(\mu))}{\mathcal{L}(\hat{\mu}, \hat{\theta})} \quad . \tag{6.20}$$

Large q values are sign that data are in bad agreement with the considered hypothesis, on the contrary small values indicate a good agreement with the hypothesis. For the purpose of establishing an upper limit on the strength parameter μ we can define the test statistic

$$q_\mu = \begin{cases} -2 \cdot \log \frac{\mathcal{L}(\mu, \hat{\hat{\theta}}(\mu))}{\mathcal{L}(\hat{\mu}, \hat{\theta})} & \hat{\mu} \leqslant \mu \\ 0 & \hat{\mu} > \mu \end{cases} \tag{6.21}$$

that is identical to the first one, but with an additional requirement that allows to take into account of only downward fluctuations of the number of signal events. The reason of introducing the constraint for $\hat{\mu} > \mu$ is that when setting an upper limit, data with $\hat{\mu} > \mu$ would not be regarded as representing less compatibility with μ than the data obtained, and therefore this is not taken as part of the rejection region of the test. The third test statistic analysed (recommended by the ATLAS Collaboration) is defined as

$$\tilde{q}_\mu = \begin{cases} -2 \cdot \log \frac{\mathcal{L}(\mu, \hat{\hat{\theta}}(\mu))}{\mathcal{L}(0, \hat{\theta}(0))} & \hat{\mu} < 0 \\ -2 \cdot \log \frac{\mathcal{L}(\mu, \hat{\hat{\theta}}(\mu))}{\mathcal{L}(\hat{\mu}, \hat{\theta})} & 0 \leqslant \hat{\mu} \leqslant \mu \\ 0 & \hat{\mu} > \mu \end{cases} \quad , \tag{6.22}$$

where, to avoid issues with negative p.d.f.s when $\hat{\mu} < 0$, the free fit likelihood term is replaced by the likelihood obtained in a fit to the zero-signal hypothesis $\mathcal{L}(0, \hat{\theta}(0))$.

The upper limit procedure adopted for all the tests reported in the next sections is based on toy MC experiments. Figure 6.19 shows an example of sampling distributions of a test statistic q for *signal+background* and *background-only* hypotheses obtained with toys MC experiments. The value of the test statistic q_{obs} (obtained evaluating

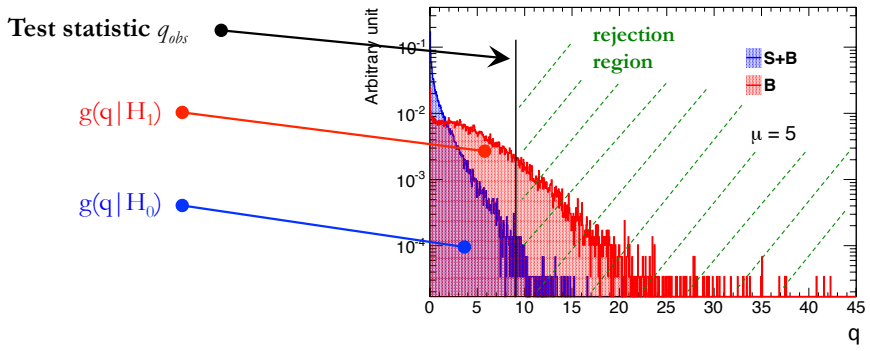


Fig. 6.19. *Example of test statistic sampling distributions for signal + background hypothesis (blue) and background-only hypothesis (red) when setting upper limits. The test statistic q_{obs} (obtained evaluating the considered test statistic q on data) defines the decision boundary which divides the space into the rejection and acceptance region of the H_0 hypothesis.*

the considered test statistic q on data) defines the decision boundary which divides the space into the rejection and acceptance region of the H_0 hypothesis.

6.3.1 The CL$_s$ Recipe in a Nut Shell

This section shows the strategy used to find the observed and expected upper limit whose computations is based on toy MC experiments. Observed data are used to define the generated signal and background models, and to evaluate the test statistic (q_{obs}) to estimate the observed upper limit. The procedure can be summarised in four main steps.

1. Construct the likelihood for signal (e.g. equation 6.11) and background (e.g. equation 6.12)
2. Choose the test statistic (6.20 or 6.21 or 6.22)
3. Perform a scan on the signal strength μ ($\mu > 0$), and for each μ:
   (a) Make toy MC experiments for H_1 hypothesis (*background-only*)
   - Generate mock data with $N_s = 0$
   - For each toy perform the fit with all parameters floating and extract the NLL $\rightarrow$ denominator of $\lambda(\mu)$
   - For each toy perform the fit fixing μ to the value of the scan and extract the NLL (NLL(μ)) $\rightarrow$ numerator of $\lambda(\mu)$
   - Evaluate the profile likelihood ratio $\lambda(\mu) = -(\text{NLL}(\mu) - \text{NLL})$

   (b) Make toy MC experiments for H_0 hypothesis (*signal+background*)
   - Generate mock data with $\mu \cdot N_s + N_b$
   - For each toy perform the fit with all parameters floating and extract the NLL $\rightarrow$ denominator of $\lambda(\mu)$

- For each toy perform the fit fixing μ to the value of the scan and extract the NLL ($\text{NLL}(\mu)$) $\rightarrow$ numerator of $\lambda(\mu)$
- Evaluate the profile likelihood ratio $\lambda(\mu) = -(\text{NLL}(\mu) - \text{NLL})$

(c) Save both the hypothesis test results for each scanned μ. The data test statistic $q_{obs}(\lambda)$ defines the rejection region of the test and it is obtained evaluating the adopted test statistic (based on the profile likelihood) on *real data* for all the different μ of the scan

4. Evaluate the CL_s

Going in more details into the last point, for each value μ of the scan we can evaluate the hypothesis test distribution $f(\tilde{q}_\mu|\mu, \hat{\hat{\theta}}(\mu, obs))$ for *signal+background* obtained from toys, the hypothesis test distribution for *background-only* obtained from toys $f(\tilde{q}_\mu|0, \hat{\hat{\theta}}(0, obs))$ and the test statistic value obtained from data $\tilde{q}_{\mu,obs}$.

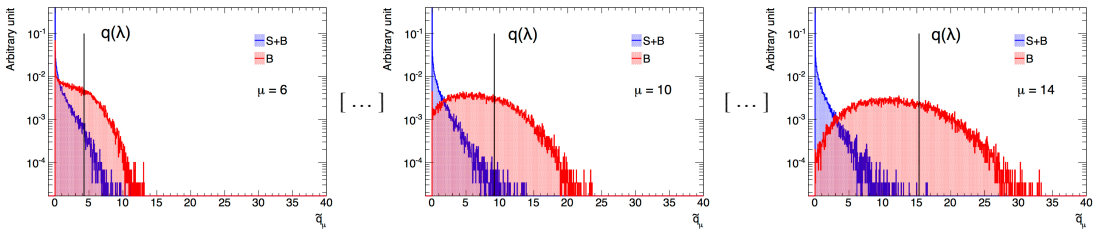


Fig. 6.20. *Example of scan on the signal strength μ.*

With these quantities we can now compute the probability, under the assumption of *signal+background* hypothesis ($s + b$), to get a value of $\tilde{q}$ with equal or lesser compatibility with the $s + b$ model relative to what is found with $\tilde{q}_{\mu,obs}$: this is called the CL_{s+b}

$$\text{CL}_{s+b} = \text{P}(\tilde{q} \geqslant \tilde{q}_{\mu,obs}|s+b) = \int_{\tilde{q}_{\mu,obs}}^{\infty} f(\tilde{q}_\mu|\mu, \hat{\hat{\theta}}(\mu, obs))\, d\tilde{q}_\mu \quad , \tag{6.23}$$

and the probability to get a value of $\tilde{q}$ with equal or higher compatibility with the *background-only* hypothesis than the observed one, called CL_b

$$\text{CL}_b = \text{P}(\tilde{q} \geqslant \tilde{q}_{\mu,obs}|b) = \int_{\tilde{q}_{\mu,obs}}^{\infty} f(\tilde{q}_\mu|0, \hat{\hat{\theta}}(0, obs))\, d\tilde{q}_\mu = 1 - p_b \quad , \tag{6.24}$$

where p_b is the p-value of the *background-only* hypothesis. Now we have all the ingredients to evaluate the CL_s

$$\text{CL}_\text{s} = \frac{\text{CL}_{s+b}}{\text{CL}_b} = \frac{p_{s+b}}{1 - p_b} \quad . \tag{6.25}$$

To cope with low sensitivity, the CL_s procedure corrects the p-value of the $s + b$ hypothesis by a number that is smaller than 1. If the hypothesis test distributions

are very well separated then p_b will be very small and $1 - p_b \sim 1$, in these conditions $\mathrm{CL}_s \sim \mathrm{CL}_{s+b}$ which is just the ordinary p-value of the *signal+background* hypothesis. On the contrary, if the two distributions are very close then the p_b will be large and $1 - p_b$ small. In this case the CL_s is prevented to become very small. The rejection is decided on the corrected p-value that prevents to reject hypotheses where there is very little sensitivity (spurious exclusion). The expected limit and bands of the well-known "Brazilian flag" plot are obtained by replacing data test statistic values with quantiles of the *background-only* test statistic distribution. The expected (median) upper limit, as well as the 1-σ (green) and 2-σ (yellow) bands are defined as:

- **Expected limit** (median) at 95 % CL → (50 % quantile)
- **1-$\boldsymbol{\sigma}$** up and down bands → 15.9 % and 84.1 % quantiles
- **2-$\boldsymbol{\sigma}$** up and down bands → 2.1 % and 97.9 % quantiles

6.3.2 Treatment of the Statistical Uncertainties

Statistical uncertainties can be taken into account in the extended maximum likelihood fit including Poisson components. In addition, if there is knowledge on the true value of a fit parameter $\boldsymbol{q}$, we can incorporate this knowledge into the fit procedure, constraining the likelihood with the nuisance parameter to take into account the systematic uncertainties. The priors for these uncertainties are just multiplied to the likelihood as:

$$\mathcal{L}(\boldsymbol{x};\boldsymbol{p},\boldsymbol{q})_{syst} = \mathcal{L}(\boldsymbol{x};\boldsymbol{p},\boldsymbol{q}) \cdot G(\boldsymbol{q}_{obs}|\boldsymbol{q},\sigma_{\boldsymbol{q}}) \quad , \tag{6.26}$$

where the p.d.f. G, used to model the uncertainties can be a Gaussian, log-normal [135] or gamma [135]. If the likelihood depends not only on the fit parameters, but also on parameters that are fixed in the fit, one may want to propagate their errors into the fit result. This can be done leaving the previously fixed parameters floating and introducing in the likelihood a G constraint term for each of them.

For example, as we have already seen in chapter 5, the branching ratio for $B_s^0 \to \mu^+\mu^-$ channel can be expressed in terms of the single event sensitivity as

$$\mathrm{BR}(B_s^0 \to \mu^+\mu^-) = \mathrm{SES} \cdot N_{\mu\mu} \tag{6.27}$$

where the SES can be considered as an efficiency correction, with uncertainty, to the event yield. Defining $\varepsilon = 1/\mathrm{SES}$ we can extend the likelihood, described in equations 6.8 and 6.9, introducing this correction

$$\begin{aligned} \mathcal{L}(\boldsymbol{m};\alpha) \quad = \quad & \mathrm{Poisson}(N_{TOT}^{obs}|\varepsilon \cdot BR + N_b) \\ & \cdot \prod_{i \in data} \left[\tfrac{\varepsilon \cdot BR}{\varepsilon \cdot BR + N_b} \cdot \mathcal{S}(\boldsymbol{m_i}) \; + \; \tfrac{N_b}{\varepsilon \cdot BR + N_b} \cdot \mathcal{B}(\boldsymbol{m_i};\alpha) \right] \\ & \cdot \mathrm{Gauss}(\varepsilon_{obs};\varepsilon,\sigma_\varepsilon) \end{aligned} \tag{6.28}$$

The procedure to evaluate the CL_s is exactly the same as the one reported above apart from the fact that when mock data are generated, ε is floated as a Gaussian[4] in each toy, while the data are fitted using 6.28 with the Gaussian constraint (implemented in `RooFit` using the option `Constrain()`). In this configuration the fit parameters are ε and BR. Other constraints can be added to the model in a similar way, e.g. on the B_s^0 mass measurement, signal Gaussian σ, etc.

6.3.3 CL_s tests on 4.9 fb^{-1} of data

The whole procedure explained in the previous sections has been entirely implemented from scratch using the `RooFit` toolkit (officially, ATLAS implements the CL_s by means of the tools contained in `RooStat`). Tests were performed on both blinded and un-blinded 2011 data.

6.3.3.1 BDT > 0.118 (2011 analysis)

For the first series of tests on the blinded dataset, toys have been generated using an exponential background model with the shape extrapolated from the fit of the sidebands without applying the BDT selection (see figure 6.21). The total number of background events in the mass region [4766,5966] MeV is obtained extrapolating the number of events from the mass fit and then rescaling it using the BDT distribution of the sidebands:

$$N_{bkg}^{SB+SR}(BDT > 0.118) = N_{bkg}^{SB+SR}(\text{fit no BDT selection}) \cdot \frac{N_{bkg}^{SB}(BDT > 0.118)}{N_{bkg}^{SB}(\text{no BDT selection})} \quad , \tag{6.29}$$

where SB and SR denote the sidebands and the signal region respectively (this allows to be less sensitive to statistical fluctuations in the number of events for higher BDT values). The number of events selected by a certain BDT value is extracted using the same procedure reported in section 6.2.5, i.e. fitting the right tail of the BDT output using a combination of exponential and hyperbola functions (the results of the fit are reported in figure 6.21). The number of background events obtained with this evaluation is 47.

The three different test statistic (q_μ^{basic}, q_μ and $\tilde{q}_\mu$) have been analysed in order to quantify their performances. The results for the expected CL_s median, obtained running 30,000 toy MC experiments, are reported in figure 6.22 and in table 6.6. The q_μ and $\tilde{q}_\mu$ return very similar values, while the expected limit computed with q_μ^{basic} is slightly different (due to the absence of the constraint on $\hat{\mu} > \mu$). Following the ATLAS recommendation, in the next tests we will use the $\tilde{q}_\mu$ as the default test statistic.

A second series of tests was performed on the un-blinded dataset to extract the observed upper limit generating toy MC experiments using both an exponential and

[4]The mean and sigma of the Gaussian are taken from the measured central value and uncertainty of the SES reported in table 5.7.

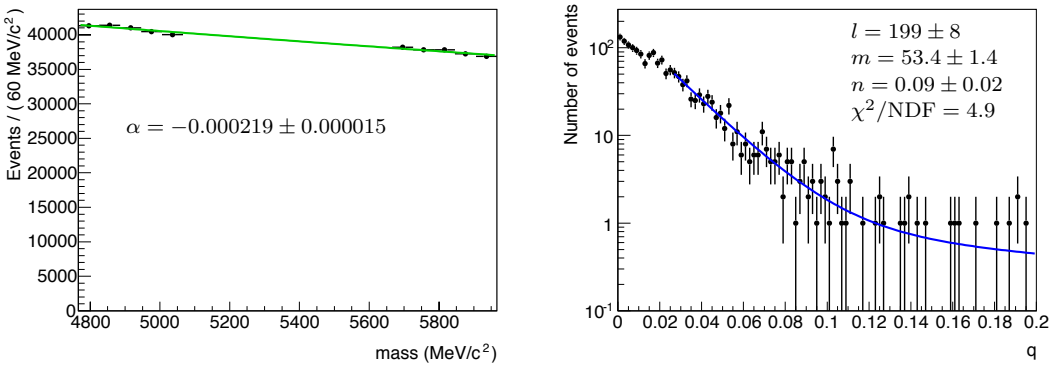


Fig. 6.21. *Invariant mass fit of the 2011 sideband data without BDT selection applied (left) and fit of the right tail of the sideband BDT distribution (right). The parameter α is the slope of the combinatorial background whereas l, m and n are the values obtained by the fit of the tail of the BDT distribution using the function reported in equation 6.16.*

Test Statistic	CL$_s$ exp.	-1σ	$+1\sigma$	-2σ	$+2\sigma$
q_μ^{basic}	2.24	1.71	3.08	1.26	4.01
q_μ	1.92	1.39	2.72	1.01	3.69
$\tilde{q}_\mu$	1.91	1.38	2.70	1.00	3.68

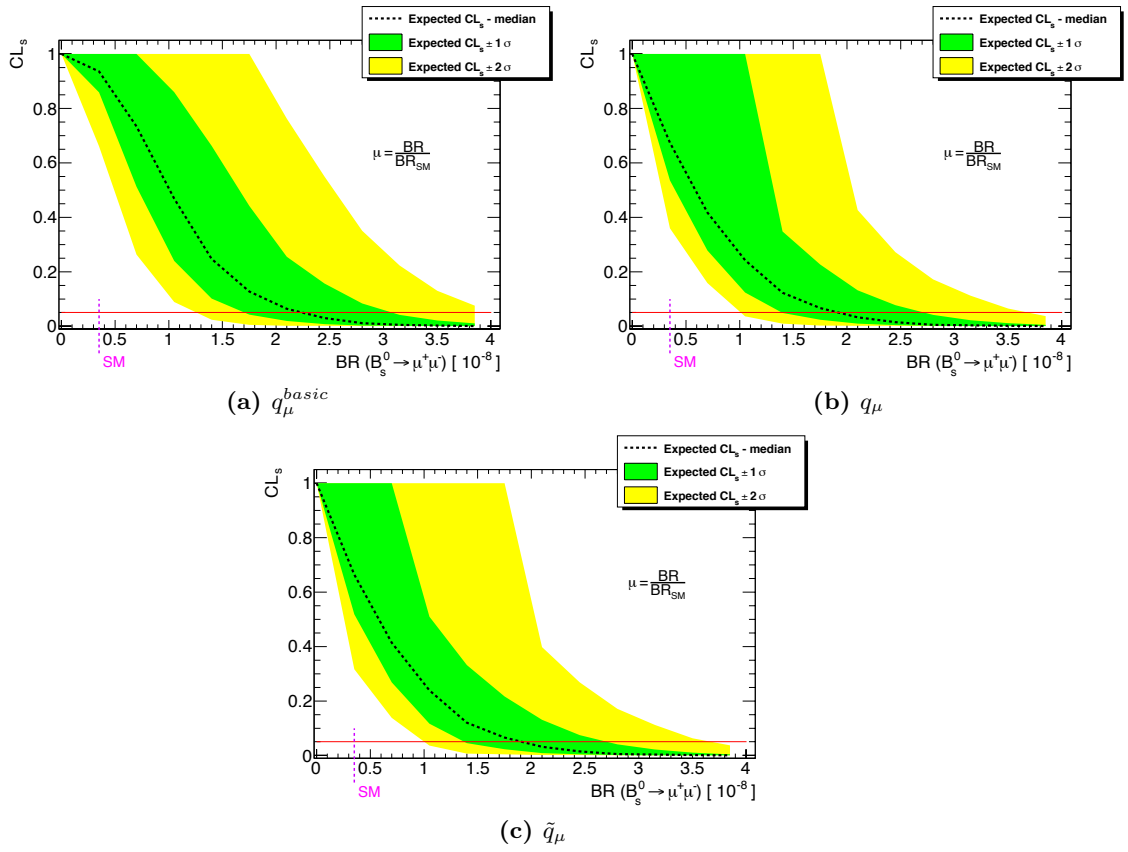
Table 6.6. *Expected (median) upper limit obtained with the three different test statistic using 2011 sideband data.*

a linear[5] background mass shape extrapolated from the fit of the un-blinded data applying the optimised BDT selection for the 2011 analysis ($BDT > 0.118$). The results of the mass fit of the background components are reported in figure 6.23.

Figure 6.24 shows the results of the mass fit (*signal+background*) and the related CL$_s$ scan obtained modelling the background with both an exponential and a linear function. The values of the expected and observed upper limit, as well as the results of the invariant mass fit, are detailed in table 6.7. The upper limit obtained with both the invariant mass fits is consistent with the one obtained with the counting procedure. When the background is modelled with an exponential a result of 0.89 ± 2.91 signal events leads to an observed upper limit higher than the expected one, while considering a linear function the under fluctuation of -3.3 events gives a lower upper limit with respect to the median one.

In order to investigate in more details the differences between counting and mass fit procedures with the two background models, the linear and the exponential fits to the background have been compared with the linear interpolation used in the 2011

[5]Beside testing the dependence of the upper limit on the background modelling, the linear function allows also to have a result more directly comparable to the one obtained with the linear interpolation used in the 2011 analysis. Nonetheless, we now have physical reasons to use this model to shape the combinatorial contribution as it will be more clear in chapters 7 and 8.



(a) q_μ^{basic} **(b)** q_μ

(c) $\tilde{q}_\mu$

Fig. 6.22. *Expected (median) CL_s as a function of $BR(B_s^0 \to \mu^+\mu^-)$ for the three different test statistic obtained using the 2011 sideband data. The number of background events generated in the toy MC experiments is 47. The horizontal red line corresponds to a CL_s equal to 0.05.*

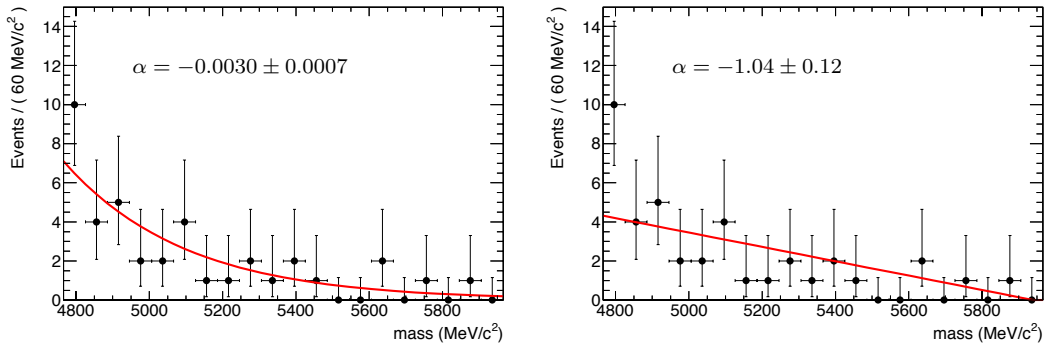
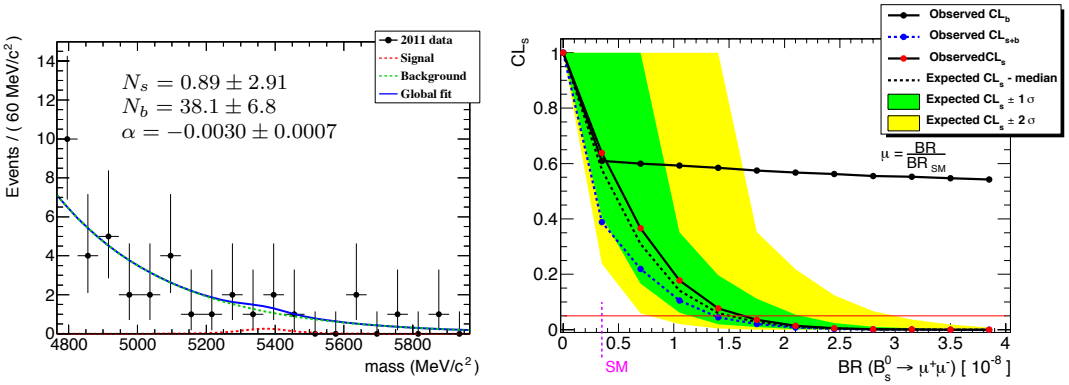
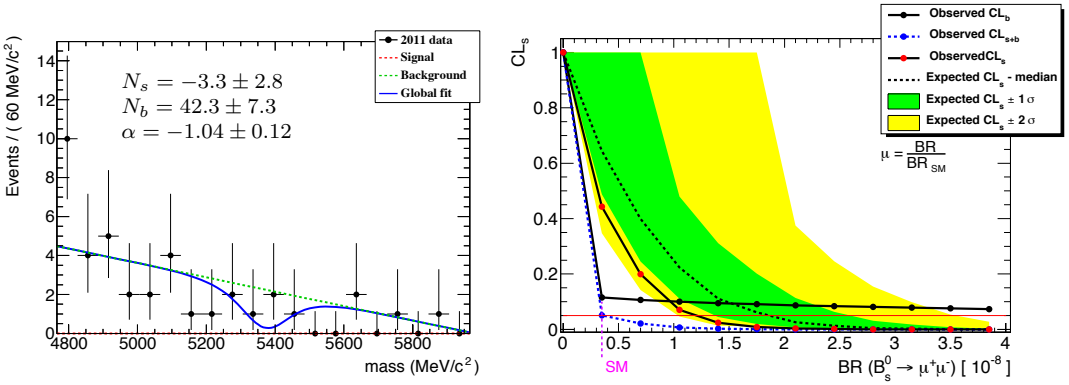


Fig. 6.23. *Invariant mass fit of the combinatorial background on 2011 un-blinded data using an exponential (left) and a linear function (right) applying the optimised BDT selection for the 2011 analysis (BDT > 0.118). The parameter α is the exponential/linear slope of the combinatorial background.*



(a) *Exponential background*



(b) *Linear background*

Fig. 6.24. *Invariant mass fit (left) and CL_s as a function of $BR(B_s^0 \to \mu^+\mu^-)$ (right) obtained modelling the combinatorial background using an exponential (**a**) and a linear (**b**) function, applying the optimised BDT selection for the 2011 analysis (BDT > 0.118). N_s and N_b are the number of signal and background events while the parameter α is the exponential/linear slope of the combinatorial background. The horizontal red line corresponds to a CL_s equal to 0.05.*

Background model	N_s	N_b
Exponential	0.89 ± 2.91	38.1 ± 6.8
Linear	-3.3 ± 2.8	42.3 ± 7.3

Background model	**CL_s exp.**	**-1σ**	**$+1\sigma$**	**-2σ**	**$+2\sigma$**	**CL_s obs.**
Exponential	1.51	1.15	2.15	0.78	3.00	1.63
Linear	1.86	1.37	2.60	1.04	3.56	1.21

Table 6.7. *Invariant mass fit yields and CL_s results obtained on 2011 un-blinded data for both exponential and linear background models applying the optimised BDT selection for the 2011 analysis (BDT > 0.118).*

analysis. The mass distributions reported in figure 6.25 allow to better understand the differences present in the values of the expected upper limit. The number of background events extracted with the linear mass fit are higher than the ones obtained by the linear interpolation, leading to a higher median expected limit ($1.86 \cdot 10^{-8}$) with respect to the one obtained in the 2011 analysis ($1.6 \cdot 10^{-8}$). On the contrary, the number of background events in the signal region extracted with the exponential are lower with respect to the ones of the 2011 analysis and, consequently, the median limit decreases. Although the differences are not huge, all these tests allow to understand the sensitivity of the upper limit to the background, and underline how a precise knowledge of the shape of the background component(s) is mandatory to obtain precise measurements (see the background studies performed for the 2012 analysis reported in chapter 8).

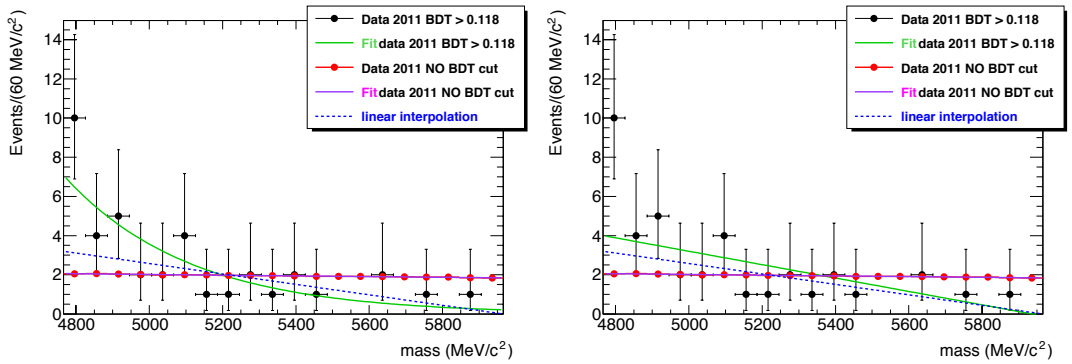


Fig. 6.25. *Comparison of the linear interpolation (blue dashed lines) with the exponential (left) and linear (right) background mass fit (green solid lines) on 2011 un-blinded data for BDT > 0.118. For shape comparison, the un-blinded dataset with no BDT selection applied (red dots), rescaled to the number of events obtained for BDT > 0.118, is reported with the related exponential fit (violet solid lines).*

6.3.3.2 BDT > 0.08 (Alternative BDT Selection Optimisation)

The same studies shown in the previous section have been performed using the BDT selection value obtained from the fit optimisation on the MC reported in section 6.2.5. For the tests on the blinded dataset, the exponential shape is extracted from the sidebands without any BDT selection, while the expected number of background events is rescaled using the fit of the BDT tail, as reported in the previous section (the number of background events obtained is 120). For the un-blinded data, the shape and the number of background events have been extracted performing the fit with both an exponential and a linear function (figure 6.26).

Since the SES depends on the applied selection, and therefore on the BDT selection considered, for the tests reported in this section a new SES should be estimated.

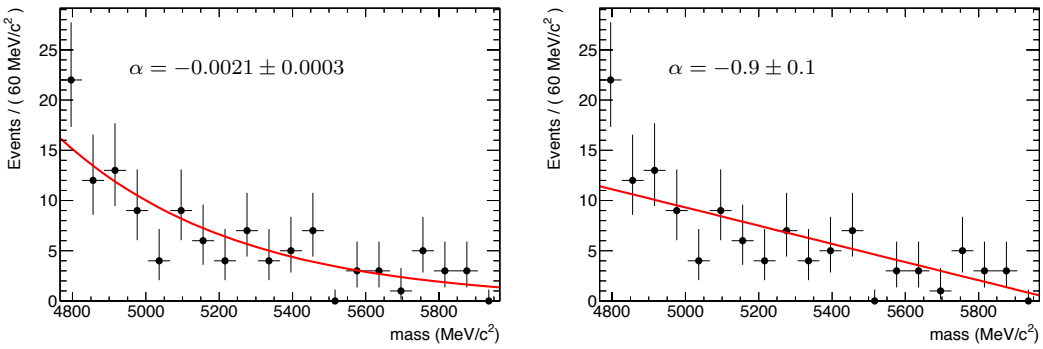


Fig. 6.26. *Invariant mass fit of the combinatorial background on 2011 un-blinded data using an exponential (left) and a linear function (right) applying the BDT selection optimised on MC (BDT > 0.08). The parameter α is the exponential/linear slope of the combinatorial background.*

Referring to chapter 5 for a more detailed explanation, we recall that the SES,

$$\mathrm{SES} = \mathrm{BR}(B^{\pm} \to J/\psi K^{\pm}) \cdot \frac{f_u}{f_s} \cdot \frac{(A_{J/\psi K^{\pm}} \cdot \varepsilon_{J/\psi K^{\pm}})}{(A_{\mu\mu} \cdot \varepsilon_{\mu\mu})} \cdot \frac{1}{N_{J/\psi K^{\pm}}} \quad , \tag{6.30}$$

can be subdivided into three main parts. The first one is composed of the $B^{\pm}$ branching ratio and the production factor. This part does not depend on the selection applied and the values are taken from the PDG and the average of the most updated measurements respectively. The second one is the efficiency times acceptance ratio, while finally the third one is the $B^{\pm}$ yield. Since the intent of these tests is only to check the CL_s and provide a rough estimation of its performance, instead of repeating the whole procedure to evaluate the efficiency times acceptance ratio,

$$\rho = \frac{(A_{J/\psi K^{\pm}} \cdot \varepsilon_{J/\psi K^{\pm}})}{(A_{\mu\mu} \cdot \varepsilon_{\mu\mu})} \quad , \tag{6.31}$$

for the given selection, the value measured in the 2011 analysis ($\rho\,(\mathrm{BDT} > 0.118) = 0.267 \pm 1.8\,[\%]\,(stat) \pm 1.1\,[\%]\,(syst)$) has been rescaled using the factor

$$K = \frac{B_s^{MC}(BDT > 0.118)}{B_p^{MC}(BDT > 0.118)} \cdot \frac{B_p^{MC}(BDT > 0.08)}{B_s^{MC}(BDT > 0.08)} = 1.018 \pm 0.009 \quad , \tag{6.32}$$

that considers the number of events in the signal MC B_s^0 (B_s^{MC}) and B^+ (B_p^{MC}) selected by the two BDT output values. The value obtained is

$$\rho\,(\mathrm{BDT} > 0.08) = K \cdot \rho\,(\mathrm{BDT} > 0.118) = 0.272 \pm 0.006 \quad . \tag{6.33}$$

The last missing ingredient in the estimation of the SES is the $B^{\pm}$ yield that has been evaluated performing a (1-dimensional) un-binned maximum likelihood fit, modelling the signal with a single Gaussian, the partially reconstructed decays using an error-function, the $B^{\pm} \to J/\psi\,\pi^{\pm}$ with a Crystal Ball, and the combinatorial background

with an exponential (more details about the fit models can be found in section 5.7). The fit, reported in figure 6.27, gives a $B^{\pm}$ yield of 21571 ± 168 events. The new SES is $(1.46 \pm 0.21) \cdot 10^{-9}$, where the systematic error of the 2011 analysis (10.4 %) was doubled to account for the not fully precise evaluation.

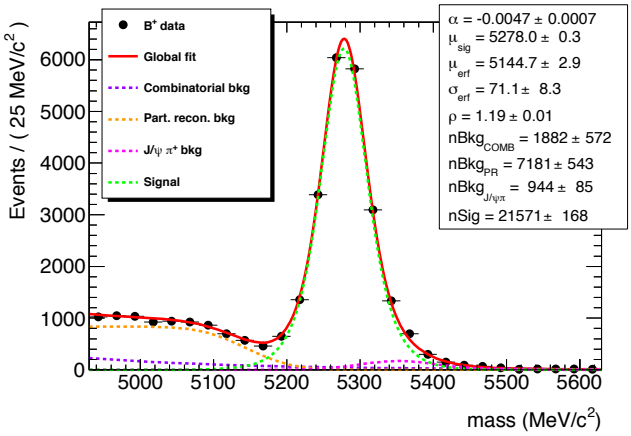


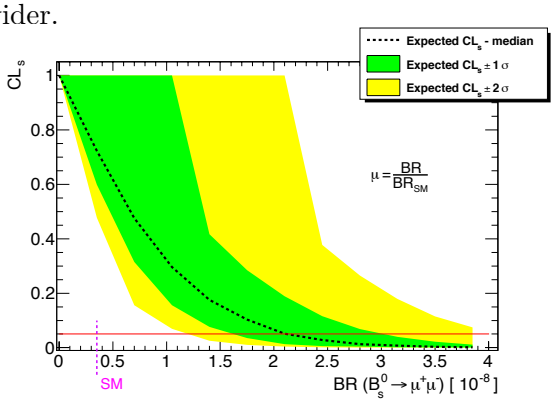
Fig. 6.27. *$B^{\pm} \to J/\psi\, K^{\pm}$ invariant mass fit for BDT > 0.08. The black dots are 2011 data, the solid red line is the total fit, the dashed green line is the signal, while the violet, orange and magenta dashed lines are respectively the combinatorial background, the partially reconstructed decays and the $B^{\pm} \to J/\psi\, \pi^{\pm}$.*

Test	BDT$_{cut}$	N_s	N_b	CL$_s$ exp.	CL$_s$ obs.
Blinded	0.118	-	47 ± 7	$1.91^{+0.79(+1.77)}_{-0.54(-0.91)}$	-
Blinded	0.08	-	120 ± 11	$2.11^{+0.91(+1.95)}_{-0.49(-0.91)}$	-
Un-blinded - exp.	0.118	0.89 ± 2.91	38 ± 7	$1.51^{+0.65(+1.49)}_{-0.36(-0.73)}$	1.62
Un-blinded - lin.	0.118	- 3.3 ± 2.8	42 ± 7	$1.86^{+0.74(+1.69)}_{-0.49(-0.83)}$	1.21
Un-blinded - exp.	0.08	1.3 ± 5.8	119 ± 12	$1.92^{+0.76(+1.81)}_{-0.52(-0.96)}$	2.06
Un-blinded - lin.	0.08	-6.5 ± 5.6	127 ± 13	$2.10^{+0.87(+1.87)}_{-0.52(-0.93)}$	1.36

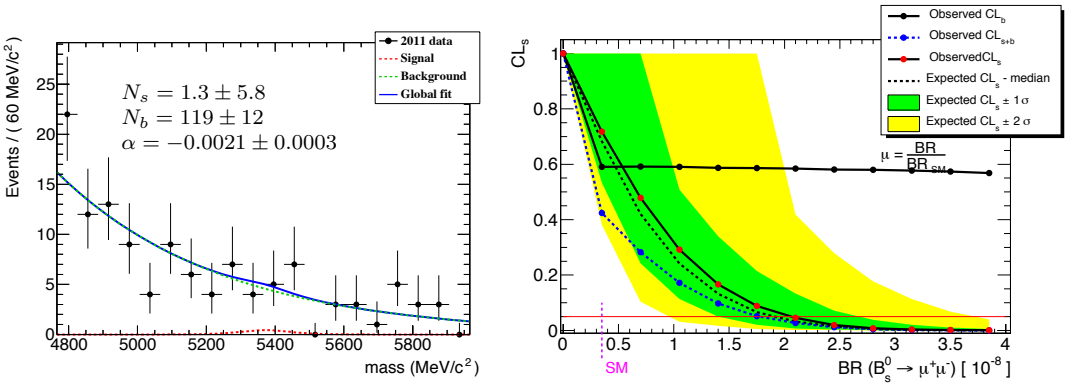
Table 6.8. *Comparison of the upper limits obtained for BDT > 0.08 and BDT > 0.118, considering both an exponential and a linear function to model the background.*

Figure 6.28 reports the results obtained on the blinded and un-blinded datasets (considering both an exponential and a linear background model) with the new BDT selection, while table 6.8 compares the upper limits with the ones obtained with the BDT selection for the 2011 analysis. Figure 6.29 reports, as in the previous section, the comparison between the exponential and the linear fit on 2011 un-blinded data with the linear interpolation obtained with the new BDT selection. With the increase

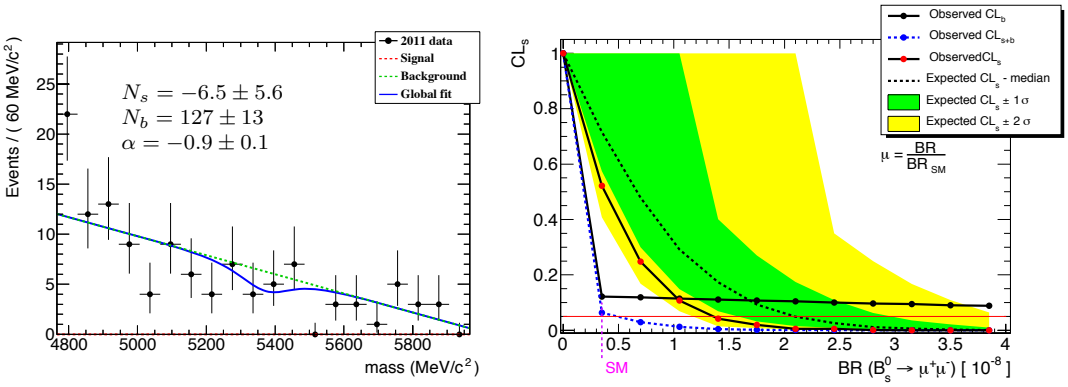
of the number of background events the difference between the linear interpolation and the mass fit is wider.



(a) *Expected CL_s obtained with the blinded dataset.*



(b) *CL_s on the un-blinded data modelling the background with an exponential function.*



(c) *CL_s on the un-blinded data modelling the background with a linear function.*

Fig. 6.28. *CL_s as a function of $BR(B_s^0 \to \mu^+\mu^-)$ for BDT > 0.08 obtained on the blinded (**a**) and un-blinded data considering both an exponential (**b**) and a linear (**c**) background model. N_s and N_b are the number of signal and background events while the parameter α is the exponential/linear slope of the combinatorial background. The horizontal red line corresponds to a CL_s equal to 0.05.*

Although the looser selection on the BDT allows to raise the number of signal

events, the gain is not enough to improve the *signal-to-noise* ratio that is half of the one obtained with the optimised BDT selection for the 2011 analysis. Due to the higher number of background events, the expected upper limit for the exponential and linear models increases of roughly a factor 1.3 and 1.1 respectively, and so does the observed upper limit.

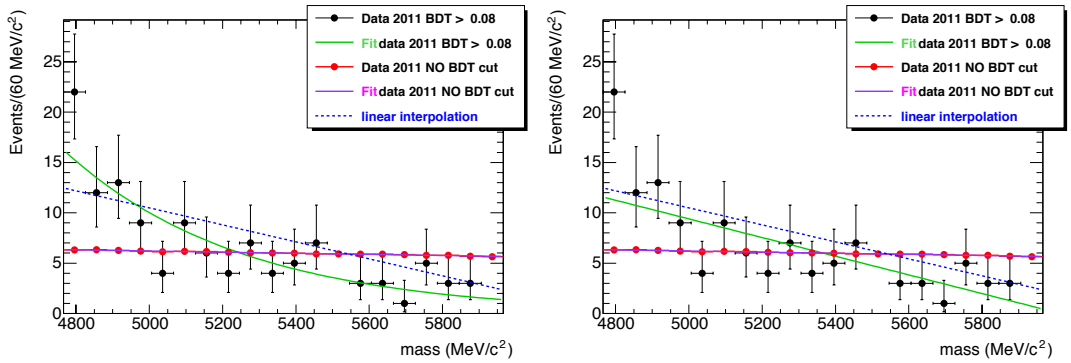


Fig. 6.29. *Comparison of the linear interpolation (blue dashed lines) with the exponential (left) and linear (right) background mass fit (green solid lines) on 2011 un-blinded data for BDT > 0.08. For shape comparison, the un-blinded dataset with no BDT selection applied (red dots), rescaled to the number of events obtained for BDT > 0.08, is reported with the related exponential fit (violet solid lines).*

All these studies show how the fit and the selection strategies have to be further optimised to guarantee an improvement in the expected signal sensitivity. The idea of performing the invariant mass fit on a broaden BDT range, trying to increase as much as possible the efficiency of the signal selection, is the key point to improve the analysis results, but has to be placed side by side to a higher separation between the signal and the background, an optimised fit procedure, and a precise and reliable modelling of all the background components.

6.3.3.3 Background systematics

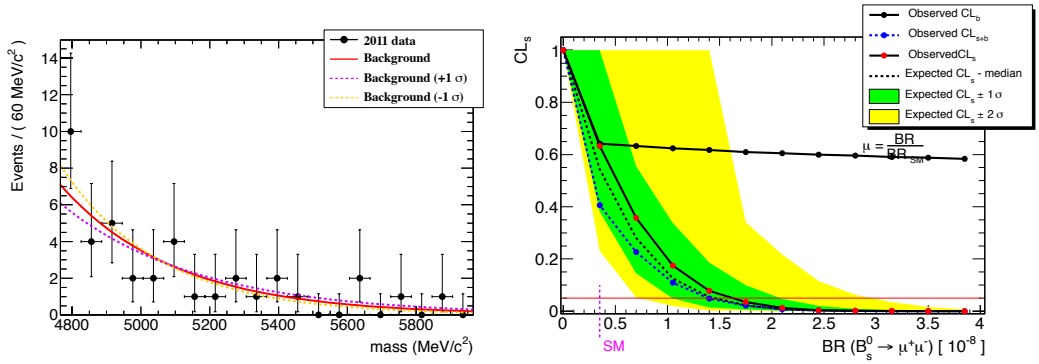
As we have already discussed above, in the estimation of the upper limit one may want to include systematic effects associated, for example, to the fit model. In this section we want to show an example of the implementation of the upper limit with toy MC experiments introducing a systematic uncertainty on the background component. As shown in section 6.3.2 this can be implemented introducing in the likelihood a Gaussian constraint on the parameter α

$$\begin{aligned}\mathcal{L}(\boldsymbol{m};\alpha) \;=\; & \text{Poisson}(N_{TOT}^{obs}|\varepsilon\cdot BR + N_b) \\ & \cdot \prod_{i\,\in\, data}\left[\frac{\varepsilon\cdot BR}{\varepsilon\cdot BR+N_b}\cdot S(\boldsymbol{m_i}) \;+\; \frac{N_b}{\varepsilon\cdot BR+N_b}\cdot B(\boldsymbol{m_i};\alpha)\right] \\ & \cdot \text{Gauss}(\varepsilon_{obs};\varepsilon,\sigma_\varepsilon) \\ & \cdot \text{Gauss}(\alpha_{obs};\alpha,\sigma_\alpha)\end{aligned} \tag{6.34}$$

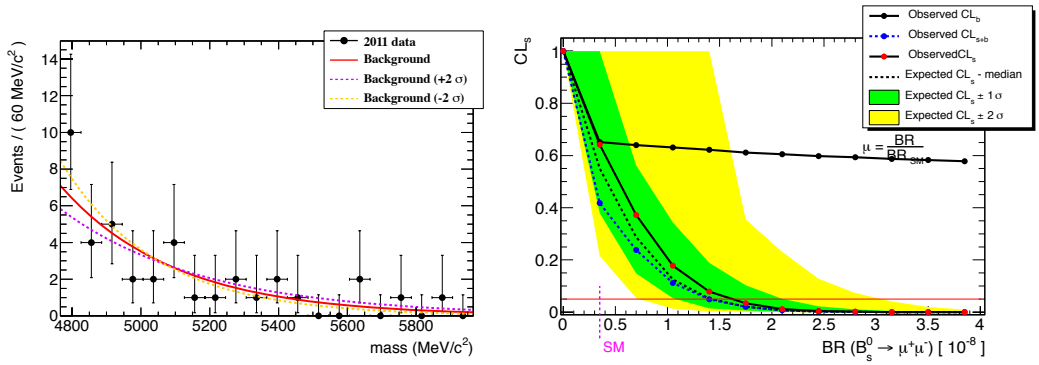
where σ_α is the systematic uncertainty associated to the background model considered. At the time when these tests were performed, a proper study on the background systematic was not finalised. In order to check the correct implementation of the method, we considered, as systematic, the statistical uncertainty (and its double value) obtained from the fit of the un-blinded data (figures 6.23 and 6.26) whose results are summarised in table 6.9.

$\mathbf{BDT}_{cut}$	Exponential slope [rel. err.]	Linear slope [rel. err.]
0.118	-0.0030 ± 0.0007 [23 %]	-1.04 ± 0.12 [12 %]
0.08	-0.0021 ± 0.0003 [15 %]	-0.9 ± 0.1 [11 %]

Table 6.9. *Summary of the background fit results on 2011 un-blinded data for exponential and linear function.*

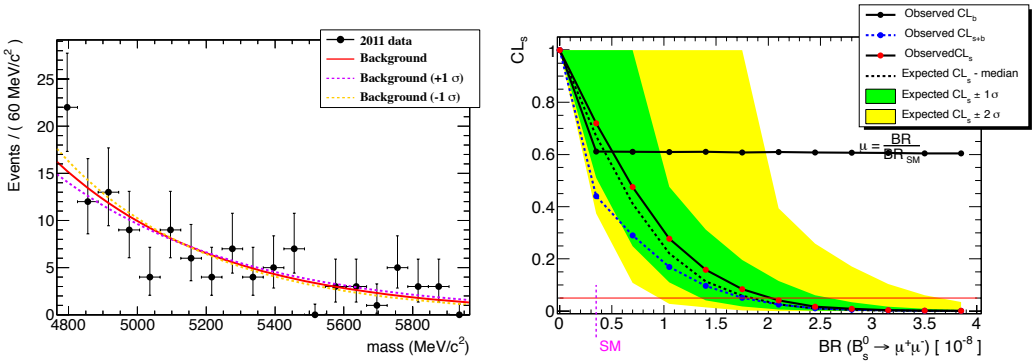


(a) *BDT > 0.118 with a constraint on the exponential equal to* $\pm 1\,\sigma$.

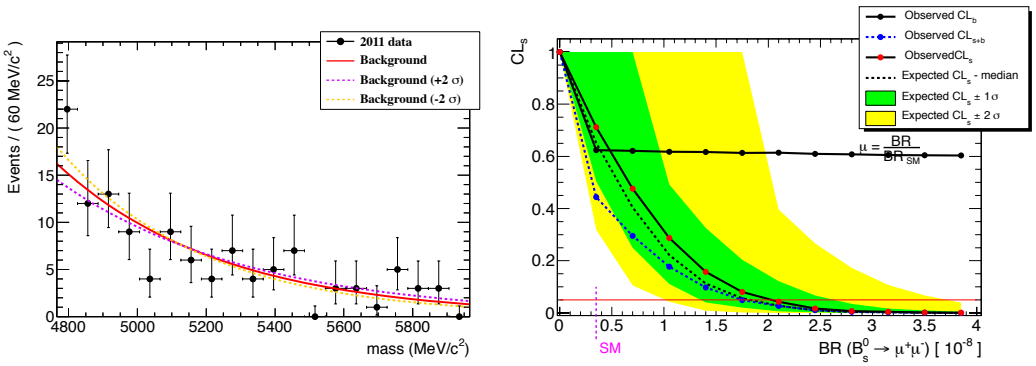


(b) *BDT > 0.118 with a constraint on the exponential equal to* $\pm 2\,\sigma$.

Fig. 6.30. *Invariant mass fit (left) and CL_s as a function of $BR(B^0_s \to \mu^+\mu^-)$ (right) obtained constraining the exponential shape in the fit $\pm 1\,\sigma$ and $\pm 2\,\sigma$ around the mean value for BDT > 0.118. The dashed yellow and violet lines in the left plot show the error on the shape extracted from the mass fit. The horizontal red line corresponds to a CL_s equal to 0.05.*



(a) *BDT > 0.08 with a constraint on the exponential equal to* $\pm 1\,\sigma$.



(b) *BDT > 0.08 with a constraint on the exponential equal to* $\pm 2\,\sigma$.

Fig. 6.31. *Invariant mass fit (left) and CL_s as a function of $BR(B^0_s \to \mu^+\mu^-)$ (right) obtained constraining the exponential shape in the fit $\pm 1\,\sigma$ and $\pm 2\,\sigma$ around the mean value for BDT > 0.08. The dashed yellow and violet lines in the left plot show the error on the shape extracted from the mass fit. The horizontal red line corresponds to a CL_s equal to 0.05.*

$\mathbf{BDT}_{cut}$	**Systematic uncertainty**	$\mathbf{CL}_s$ **exp.**	$\mathbf{CL}_s$ **obs.**
0.118	$\pm 1\,\sigma$	$1.45^{+0.64(+1.51)}_{-0.40(-0.72)}$	1.63
0.118	$\pm 2\,\sigma$	$1.42^{+0.70(+1.63)}_{-0.34(-0.69)}$	1.64
0.08	$\pm 1\,\sigma$	$1.85^{+0.75(+1.80)}_{-0.49(-0.89)}$	2.03
0.08	$\pm 2\,\sigma$	$1.83^{+0.80(+1.88)}_{-0.48(-0.82)}$	2.05

Table 6.10. *Expected and observed upper limits obtained constraining the exponential shape in the fit $\pm 1\,\sigma$ and $\pm 2\,\sigma$ around the mean value obtained from the fit on the 2011 un-blinded data.*

Figures 6.30 and 6.31 show the results obtained, for the two BDT selections adopted in the studies reported in the previous sections, constraining the background shape in

the fit $\pm 1\,\sigma$ and $\pm 2\,\sigma$ around the mean value obtained from the fit on the un-blinded data. The expected and observed upper limits obtained are summarised in table 6.10, showing to be compatible with the ones measured leaving the slope unconstrained.

Chapter 7

Studies in Preparation for the 2012 Analysis

The tests performed on toy MC experiments and on 2011 un-blinded data reported in chapter 6 showed, on one side, the stability and the reliability of the invariant mass fit, but on the other side underlined the necessity of increasing the expected sensitivity (limited in 2011 also by statistics). Therefore, in order to improve the signal event selection, raising the rejection power on the background components, additional studies on the MVA have been performed. In addition, with the intent of investigating new fit strategies (section 7.2), like the two dimensional fit on the mass and MVA classifier, the shape of the multivariate output should have a less complex structure than the one obtained in the 2011 analysis (see figure 5.22) in order to be parameterised in a easier way.

This chapter reports the studies I performed, in preparation for the 2012 analysis, on the MVA and on the possible new fit strategies for the extraction of the B_s^0 and B^0 event yields with the maximum significance. A comparison of the ATLAS and CMS performance will be also shown in section 7.2.4.4.

All the tests reported in this chapter are not documented in the ATLAS note, but they have been fundamental for the 2012 analysis reported in chapter 8. The fit I developed for these studies is used to extract the signal yield in 2012 data.

7.1 MVA Studies

Different alternatives to the BDT classifier adopted for the 2011 analysis have been investigated, and among all of them the Fisher estimator seems to be very promising to develop the new fit strategies analysed in the next sections since, by construction, it has a Gaussian-like shape. Nevertheless, the performance in terms of signal-background separation should be carefully studied.

7.1.1 The Fisher Discriminant

The Fisher linear discriminant [136] is a classification method that projects high-dimensional data onto a line and performs classification in this one-dimensional space. Given a set of n observations x_i, we can write the linear combination

$$t(\vec{x}) = \sum_{i=1}^{n} a_i x_i \quad , \tag{7.1}$$

where the values of the parameters a_i are chosen in order to maximise the separation between the signal $\mathcal{M}(t;s)$ and background $\mathcal{M}(t;b)$ p.d.f.s. Assuming that the two p.d.f.s are both normally distributed with mean μ_0 and μ_1, and covariance Σ_0 and Σ_1 respectively, the Fisher discriminant maximises

$$J(\vec{a}) = \frac{(\mu_1 - \mu_0)^2}{\Sigma_1^2 + \Sigma_2^2} \quad , \tag{7.2}$$

i.e. it maximises the distance between the means of the two classes (separation between classes), minimising the variance within each class (separation within the classes).

The first series of tests was performed starting from the thirteen discriminating variables used for the 2011 analysis. Since some of those (like B meson p_T and p_L^{min}), as reported in section 6.1.1, showed a higher correlation with the mass with respect to the others, various subsets have been investigated, trying to retain only variables that exhibit a small dependence on the mass, but at the same time paying carefully attention to do not lose power in terms of signal-background separation, and therefore maintaining the same performance. In addition, despite its low mass correlation, also the proper time significance has been (for the moment) removed in these tests, as it provides very similar information to the L_{xy}. The search has been restricted to the two sets summarised in table 7.1, the first contains nine variables, while the second only three.

The Fisher classifier has been trained on 2011 MC re-weighted $b\bar{b} \to \mu^+\mu^- X$ and $B_s^0 \to \mu^+\mu^-$ events, and the results obtained are reported in figure 7.1 that shows the signal and background output distributions for the two configurations reported above. The separation, $\mathbb{S}$, between two distributions D_1 and D_2 of the variable x is defined as

$$\mathbb{S} = \frac{1}{2} \int_{-\infty}^{+\infty} \frac{(D_1(x) - D_2(x))^2}{D_1(x) + D_2(x)} \, dx \quad , \tag{7.3}$$

and it has been evaluated using the TMVA [118] function `GetSeparation()`. Despite the very different number of variables present in the two configurations, the separation showed to be very similar, as underlined also by the small differences in the Fisher output shapes. Figure 7.2 reports the efficiency for signal and background, and the Receiver Operating Characteristic (ROC) curve (background rejection versus signal efficiency) for the classifier trained with three and nine variables, further proving how the performance in the two cases are very similar.

In order to understand in more detail its capabilities, the performance obtained with the Fisher classifier are compared with the ones obtained with the BDT used

	Configuration with 9 variables	Configuration with 3 variables
1	Isolation ($I_{0.7}$)	Isolation ($I_{0.7}$)
2	ΔR	ΔR
3	L_{xy}	L_{xy}
4	α_{2D}	
5	DCA	
6	ZCA	
7	d_0^{max}	
8	$\log(\chi^2(PV\text{-}SV))$ in the x-y plane	
9	$\log(\chi^2(PV\text{-}SV))$ along z axis	

Table 7.1. *List of the variables selected for the two configurations used in the MVA studies.*

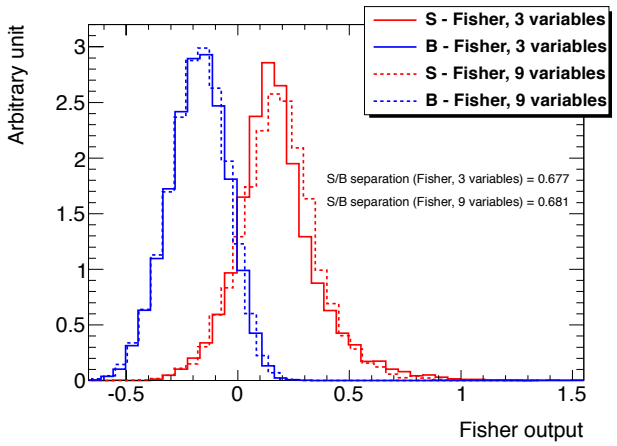


Fig. 7.1. *Fisher output distribution obtained for the two configurations with three (solid lines) and nine (dashed lines) variables. In red are reported the signal distributions, while the background is in blue.*

for the 2011 analysis. Figure 7.3 reports the efficiency, the ROC curve, as well as the Punzi [119] and the $s/\sqrt{s+b}$ estimators for both the Fisher and the BDT. The better separation provided by the BDT is evident in the ROC curve even if the difference is not big. Nonetheless, the small loss in separation power of the Fisher is acceptable with the intent of obtaining an easier parametrisable output shape, fundamental for the 2-dimensional fit implementation.

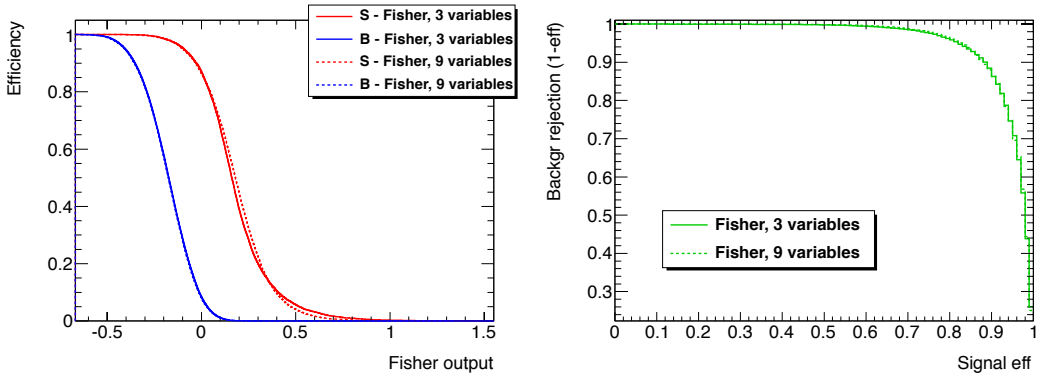


Fig. 7.2. *Signal and background efficiency as a function of the classifier output (left), and ROC curve (right) obtained with the Fisher trained on nine variables (dashed lines) and three variables (solid lines). In the left plot the signal is in red, while the background in blue.*

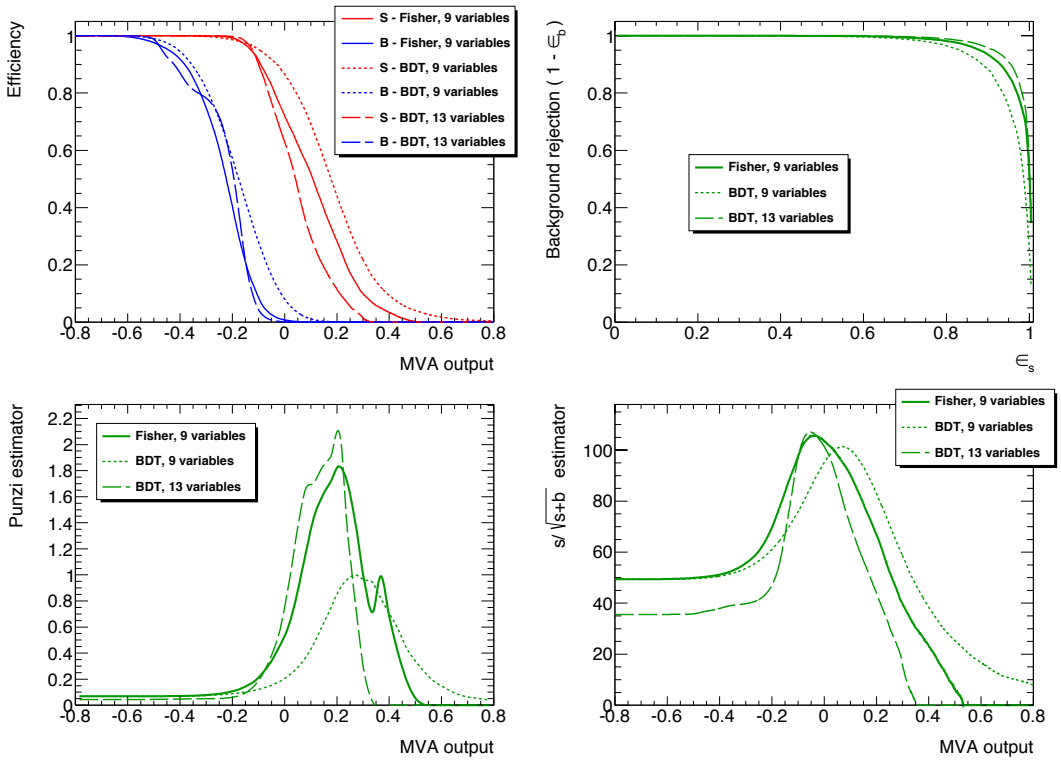
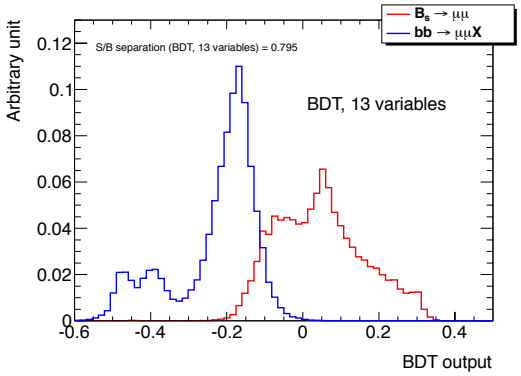


Fig. 7.3. *From left to right, top to bottom: signal and background efficiency as a function of the MVA output, background rejection versus signal efficiency, Punzi and $s/\sqrt{s+b}$ estimators as a function of the MVA output obtained with the Fisher trained on nine variables and with the BDT trained on both nine an thirteen variables.*

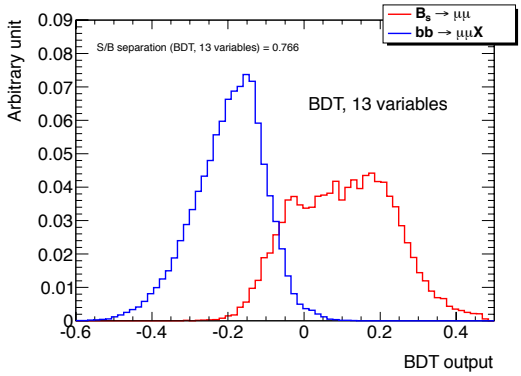
7.1.2 Effects of the Additional Cuts on the MVA Output

In section 6.1.3 we introduced some additional selection on the discriminating variables to reduce the amount of background before any MVA selection or fit. Beside

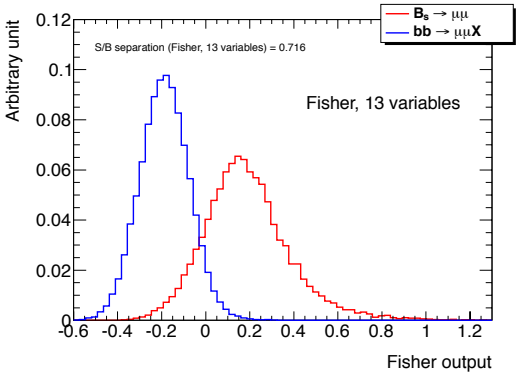
that, the selection on the ΔR and α_{2D} allow also to remove the double peak structures present in their distributions around π radians (see figure 6.8).



(a) *BDT with no additional selection.*



(b) *BDT with* $\alpha_{2D} < 1.0$, $\Delta R < 1.5$, $L_{xy} > 0$.



(c) *Fisher with* $\alpha_{2D} < 1.0$, $DR < 1.5$, $L_{xy} > 0$.

Fig. 7.4. *Effect of the additional selection on the BDT distribution. The same BDT configuration used to obtain the BDT output distribution used in the 2011 analysis (a) has been trained introducing the additional selection detailed in section 6.1.3 (b). The Fisher classifier with the additional selection applied and trained with the same variables is reported as reference.*

Figure 7.4 shows the 2011 BDT distribution compared with the one obtained performing the training after having applied the additional selection: the removal of the double structure in the background allows to regularise the output distribution. The separation between signal and background is lower in the configuration with the additional selection, but this is only an apparent effect due to the reduction of the continuum that leads to a reduced number of events in the regions of the phase-space far away from the signal. For comparison, in figure 7.4 it is also reported the Fisher output obtained with the default thirteen discriminating variables, that shows a 7 % lower separation with respect to the same configuration trained with the BDT.

7.2 New Fit Strategies

The invariant mass fit results showed in chapter 6 have been obtained applying a selection specifically optimised for a "cut & count" procedure. While removing the background in the signal region, this method cuts away also a considerable amount of signal events. In order to maximise the sensitivity of the $B_s^0 \to \mu^+\mu^-$ analysis for the 2012 data, new strategies for the signal extraction have been investigated. The first one is based on a two dimensional likelihood fit (invariant mass - MVA output), while the second one categorises the events according to different interval in the MVA output and then the mass distributions are fitted simultaneously to extract the parameters of interest. Both strategies allow to maintain a high signal efficiency.

For these two methods, a first series of tests was done modelling the background using an exponential function. However, this assumption is not completely correct in the most significant (signal-like) MVA bins in which backgrounds with different topologies with respect to the continuum have a non-negligible contribution, that could even be dominant (from now on, we will call this background "low-mass background" since it populates the low-mass part of our search region). In these cases a simple exponential is not able to reliably describe the data and thus it is replaced by a combination of a linear and an exponential function to account for the two components (see chapter 8 for more details).

Sections 7.2.1 and 7.2.2 will give a more general overview of the fit strategies that have been studied to extract the signal yield. A simple model composed of a double Gaussian for the $B_s^0 \to \mu^+\mu^-$ mass shape and one exponential for the background will be used to estimate the performance of the two fits. Then a test of the 2D fit on the $B^\pm \to J/\psi\, K^\pm$ reference channel data will prove the robustness of the procedure even in a more complex framework. In these preliminary tests reported in the following two sections, an increased signal statistics with a *signal-to-noise* ratio higher than the one expected both in 2011 and 2012 data is simulated to test the performance of the fit and especially to check the right implementation of the two methods.

After these preliminary studies, more detailed and accurate tests, based on 2011 un-blinded data and on toy MC experiments simulating the expected statistic for the 2012 analysis, will be reported in sections 7.2.3 and 7.2.4.

7.2.1 2D Invariant Mass Fit

The first fit strategy investigated is the 2D fit (invariant mass - MVA output). For these tests, the Fisher classifier has been chosen due to the easy parameterisation of its shape. The configuration used is the one with nine variables (described in section 7.1.1), applying the additional selection for the background reduction (section 6.1.3).

The signal Fisher distribution, F, has been modelled using a double Gaussian

$$\mathcal{M}_s^{\mathrm{F}}(F \,|\, \mu_s^F, \sigma_{1,s}^F, \sigma_{2,s}^F, k_s^F) = k_s^F \cdot e^{-\frac{(F-\mu_s^F)^2}{2\sigma_{1,s}^F}} + (1-k_s^F) \cdot e^{-\frac{(F-\mu_s^F)^2}{2\sigma_{2,s}^F}} \tag{7.4}$$

with the shape fixed to the MC distribution, while for the background a single Gaussian

$$\mathcal{M}_b^{\mathrm{F}}(F\,|\,\mu_b^F,\sigma_b^F) = e^{-\frac{(F-\mu_b^F)^2}{2\sigma_b^F}} \tag{7.5}$$

with the shape fixed to the 2011 sideband data, has been adopted. The results of the two fits are reported in figure 7.5. The invariant mass distribution for signal events

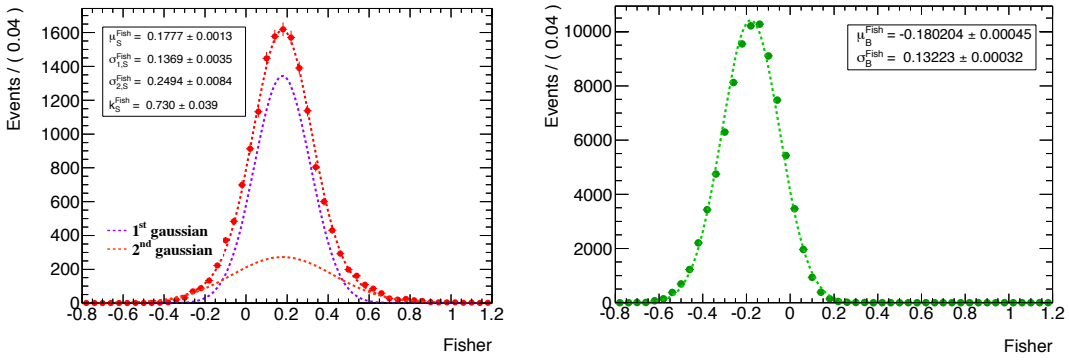


Fig. 7.5. *Fit of the Fisher output for $B_s^0 \to \mu^+\mu^-$ signal (left) and 2011 sideband data (right).*

has been modelled using a double Gaussian

$$\mathcal{M}_s^m(m\,|\,\mu_s^m,\sigma_{1,s}^m,\sigma_{2,s}^m,k_s^m) = k_s^m \cdot e^{-\frac{(m-\mu_s^m)^2}{2\sigma_{1,s}^m}} + (1-k_s^m)\cdot e^{-\frac{(m-\mu_s^m)^2}{2\sigma_{2,s}^m}} \tag{7.6}$$

with all parameters fixed to the MC shape, while the background is parameterised with an exponential

$$\mathcal{M}_b^m(m\,|\,\alpha_b) = e^{\alpha_b\cdot m} \quad , \tag{7.7}$$

where the slope used to generate the toy MC experiments is extracted from 2011 un-blinded data (applying the additional selection to reduce the background). The results of these two fits are reported in figure 7.6.

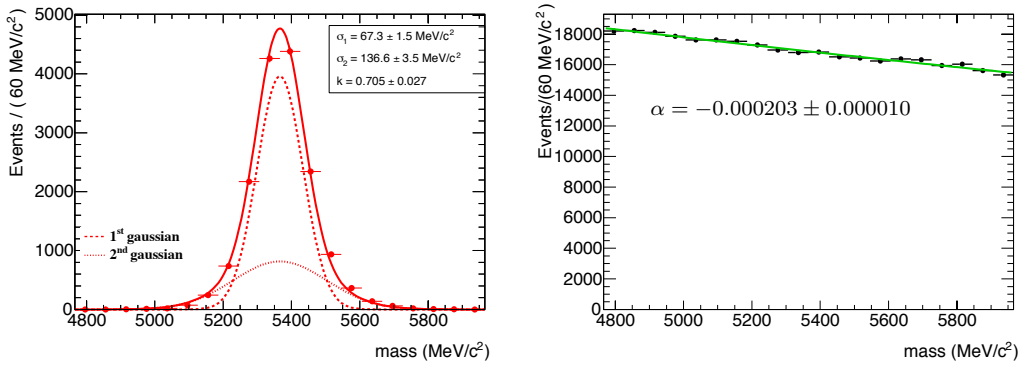


Fig. 7.6. *Fit of the invariant mass for $B_s^0 \to \mu^+\mu^-$ signal (left) and 2011 un-blinded data after having applied the additional selection for the background reduction (right).*

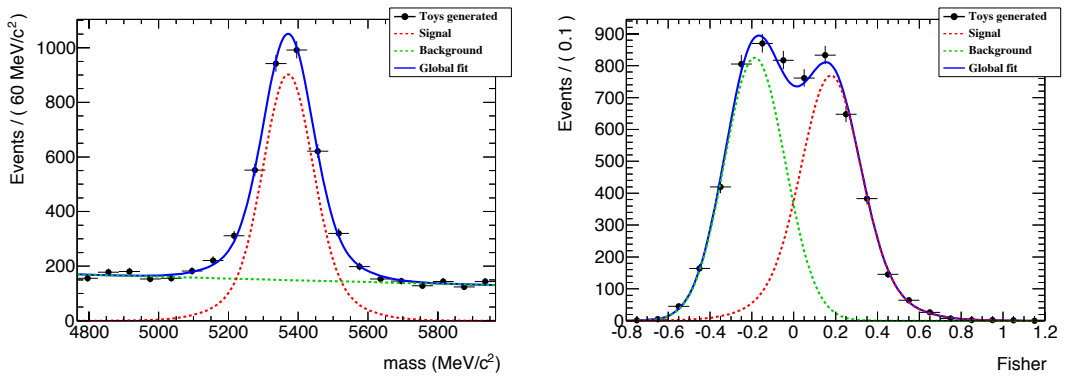
With the p.d.f.s described above we can now build the 2-dimensional models for signal (s) and background (b)

$$\begin{aligned} \mathcal{M}_s^{F,m} &= \mathcal{M}_s^{\mathrm{F}}(F\,|\,\mu_s^F, \sigma_{1,s}^F, \sigma_{2,s}^F, k_s^F) \;\times\; \mathcal{M}_s^m(m\,|\,\mu_s^m, \sigma_{1,s}^m, \sigma_{2,s}^m, k^m) \\ \mathcal{M}_b^{F,m} &= \mathcal{M}_b^{\mathrm{F}}(F\,|\,\mu_b^F, \sigma_b^F) \;\times\; \mathcal{M}_b^m(m\,|\,\alpha_b) \end{aligned} \tag{7.8}$$

and the 2-dimensional likelihood is therefore

$$\mathcal{L}_{F,m}^{2D} = \frac{N_s}{N_s + N_b} \cdot \mathcal{M}_s^{F,m} + \frac{N_b}{N_s + N_b} \cdot \mathcal{M}_b^{F,m} \quad . \tag{7.9}$$

With this model, an un-binned extended maximum likelihood fit is performed to extract all the parameters of interest. The number of signal and background events, as well as the shape of the background are free in the fit, while the shape parameters of the signal mass double Gaussian and of the Fisher (for both the signal and the background component) are fixed as already stated above. In a first attempt the invariant mass and the two Fisher shapes have been left floating, but the fit was not able to provide reliable results.



(a) *Projection in the invariant mass dimension* **(b)** *Projection in the Fisher dimension*

Fig. 7.7. *Example of 2D fit obtained generating 3,000 signal and 3,000 background events. The blue solid line is the total fit, the green dashed line is the combinatorial background and the red dashed line is the signal.*

As a proof of principle, the procedure has been tested on mock data generating 3,000 events of signal and 3,000 events of background. Figure 7.7 shows the projection of the 2D fit in the mass and Fisher dimensions obtained generating one single toy experiment, while figure 7.8 reports the distribution of the floating parameters, as well as their pulls obtained simulating 30,000 toy MC experiments (the mean values and variance of the parameters are summarised in table 7.2). The small bias on the signal (0.04 %), and the small pull of the background yield and slope show how the fit procedure is reliable and stable.

Before studying in more details the behaviour of the 2D fit in the extraction of the $B_s^0 \to \mu^+\mu^-$ signal yield, a test was made applying the fit procedure to a more

Parameter	Generated	From fit	Pull	
			Mean	Variance
N_s	3000	2999 ± 66	-0.018 ± 0.006	1.013 ± 0.004
N_{comb}	3000	3000 ± 66	-0.010 ± 0.006	1.008 ± 0.004
α_{comb}	-0.000203	-0.000203 ± 0.000001	0.004 ± 0.005	1.009 ± 0.004

Table 7.2. *Results of the toy MC experiments for the 2D fit generating 3,000 signal events and 3,000 background events.*

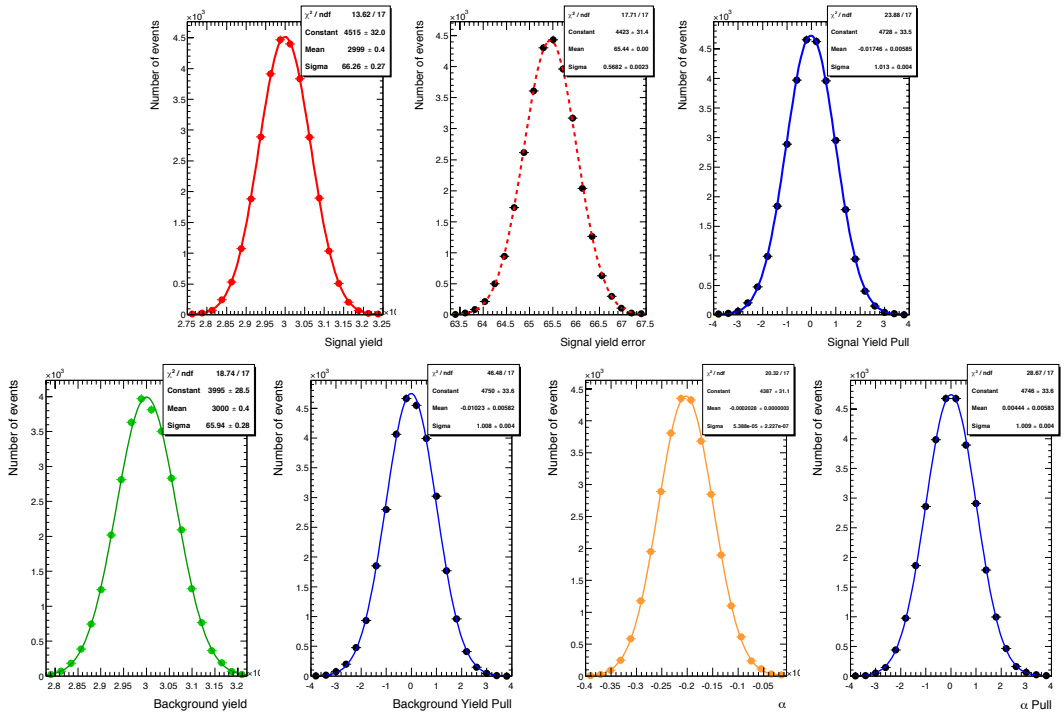


Fig. 7.8. *Mean values and pulls for the parameters of interest extracted with the 2D fit. The top plot shows the signal yield and pull, while the two bottom plots report the background yield (left) and the background slope (right) with the related pulls.*

complex model, i.e. the reference channel $B^{\pm} \to J/\psi\, K^{\pm}$. This is quite useful to test the reliability of the 2D fit in a framework in which other contributions (low-mass decays for the background and B^0 for the signal) can not be neglected.

For the mass shapes, the same models used for the fit reported in section 5.7 have been adopted. The signal has been parameterised using a single Gaussian, the combinatorial background with an exponential, the partially reconstructed decays with an error function and the $B^{\pm} \to J/\psi\, \pi^{\pm}$ with a Crystal Ball (with all parameters fixed to the MC). Unlike the fit reported in section 5.7, no per-event mass error has been used for this fit configuration.

Figure 7.9 shows the Fisher output distributions for $B^+ \to J/\psi\, K^+$ signal and for the other background components. Since the distributions of the $B^+ \to J/\psi\, \pi^+$ and partially reconstructed decays are very similar to the one of the B^+ signal (having both contributions a signal-like topology), for these preliminary tests the same model (double Gaussian with parameters fixed to 2011 MC $B^+ \to J/\psi\, K^+$) has been used to shape the two Fisher classifiers. For the combinatorial background a double Gaussian, with the parameters fixed to the fit performed on only the combinatorial component of the 2011 MC $b\bar{b} \to \mu^+\mu^- X$ (reconstructed with $B^\pm$ algorithms), is adopted.

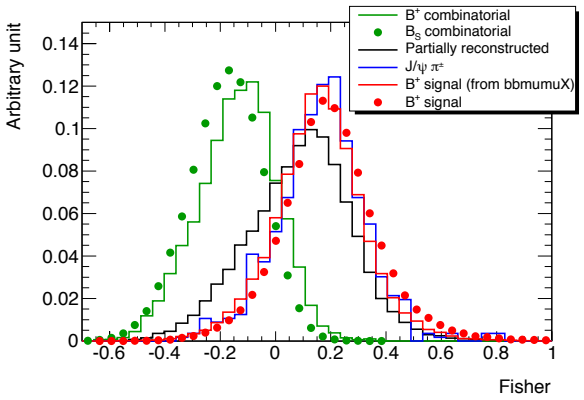


Fig. 7.9. *Fisher output distributions for the $B^+ \to J/\psi\, K^+$ signal and for the other background components.*

The 2D likelihood shown in equation 7.9 has to be modified in order to take into account the different background components:

$$\begin{aligned}
\mathcal{L}^{2D}_{F,m} &= \frac{N_s}{N_s + N_{comb} + N_{p.r.} + N_{J/\psi\pi}} \cdot \mathcal{M}^{F,m}_{s} \\
&+ \frac{N_{comb}}{N_s + N_{comb} + N_{p.r.} + N_{J/\psi\pi}} \cdot \mathcal{M}^{F,m}_{comb} \\
&+ \frac{N_{p.r.}}{N_s + N_{comb} + N_{p.r.} + N_{J/\psi\pi}} \cdot \mathcal{M}^{F,m}_{p.r.} \\
&+ \frac{N_{J/\psi\pi}}{N_s + N_{comb} + N_{p.r.} + N_{J/\psi\pi}} \cdot \mathcal{M}^{F,m}_{J/\psi\pi}
\end{aligned} \tag{7.10}$$

where

$$\begin{aligned}
\mathcal{M}^{F,m}_{s} &= \mathcal{M}^{\mathrm{F}}_{s}(F \,|\, \mu^F_s, \sigma^F_{1,s}, \sigma^F_{2,s}, k^F_s) \;\times\; \mathcal{M}^{m}_{s}(m \,|\, \mu^m_s, \sigma^m_s) \\
\mathcal{M}^{F,m}_{comb} &= \mathcal{M}^{\mathrm{F}}_{comb}(F \,|\, \mu^F_{comb}, \sigma^F_{1,comb}, \sigma^F_{2,comb}, k^F_{comb}) \;\times\; \mathcal{M}^{m}_{comb}(m \,|\, \alpha_{comb}) \\
\mathcal{M}^{F,m}_{p.r.} &= \mathcal{M}^{\mathrm{F}}_{s}(F \,|\, \mu^F_s, \sigma^F_{1,s}, \sigma^F_{2,s}, k^F_s) \;\times\; \mathcal{M}^{m}_{p.r.}(m \,|\, \mu^m_{p.r.}, \sigma^m_{p.r.}) \\
\mathcal{M}^{F,m}_{J/\psi\pi} &= \mathcal{M}^{\mathrm{F}}_{s}(F \,|\, \mu^F_s, \sigma^F_{1,s}, \sigma^F_{2,s}, k^F_s) \;\times\; \mathcal{M}^{m}_{J/\psi\pi}(m \,|\, \mu^m_{J/\psi\pi}, \sigma^m_{J/\psi\pi}, \alpha^m_{J/\psi\pi}, k^m_{J/\psi\pi})
\end{aligned} \tag{7.11}$$

The mass terms $\mathcal{M}^m_s$, $\mathcal{M}^m_{comb}$, $\mathcal{M}^m_{p.r.}$ and $\mathcal{M}^m_{J/\psi\pi}$ are described by equations 5.9, 5.10, 5.11 and 5.12 respectively, but no per-event mass error has been used in this fit configuration.

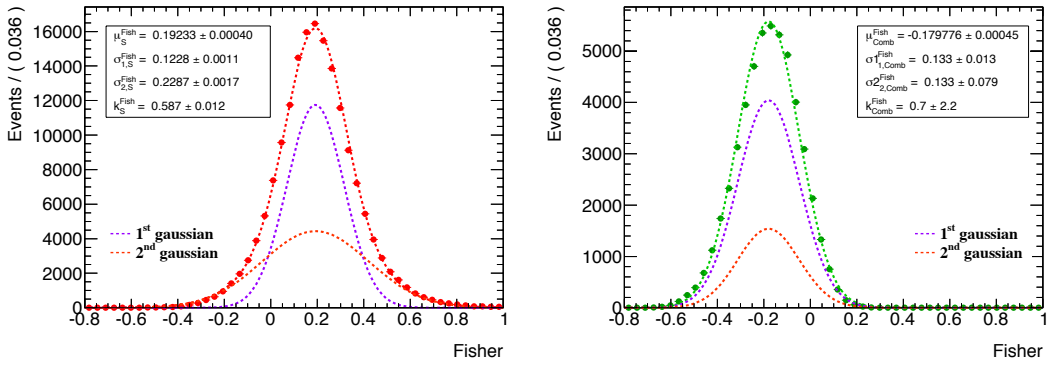


Fig. 7.10. *Fit on the Fisher output distribution for 2011 MC signal $B^+ \to J/\psi\, K^+$ (left) and combinatorial background $b\bar{b} \to \mu^+\mu^- X$ (right).*

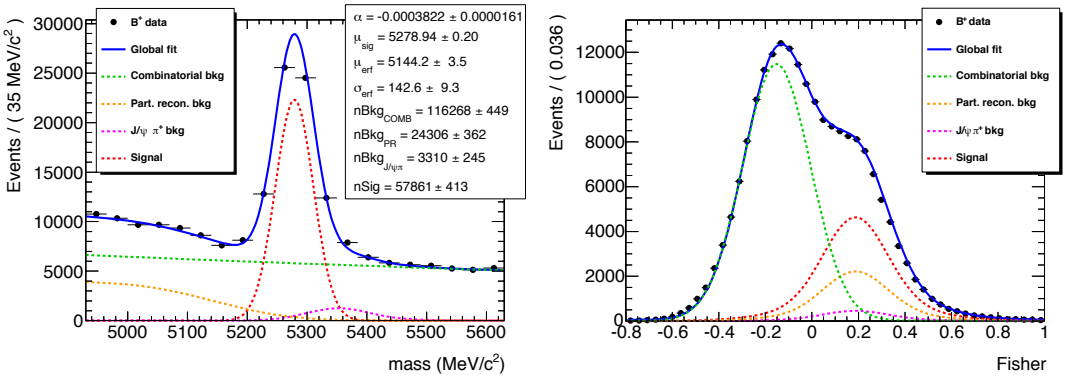


Fig. 7.11. *Projection of the 2D fit on $B^\pm$ 2011 data into the mass dimension (left) and Fisher dimension (right).*

Figure 7.10 reports the results of the fit on the Fisher distribution for 2011 MC signal $B^+ \to J/\psi\, K^+$ and combinatorial background $b\bar{b} \to \mu^+\mu^- X$, while figure 7.11 shows data projection along the mass and the Fisher dimension of the results obtained from the 2D fit. The fit on the reference channel data provides reasonable results: the signal yield extracted is roughly four times the value obtained with the 2011 analysis and therefore is fully consistent with it considering the efficiency of the 2011 BDT ($\sim 25\,\%$), while the ratio $(J/\psi\,\pi^\pm)/(J/\psi\,K^\pm)$ equal to 0.057 ± 0.004 shows a discrepancy with respect to the most updated measurement [137] (0.040 ± 0.004). It must be remembered that the Fisher distributions for the partially reconstructed decays and for the $B^+ \to J/\psi\,\pi^+$ have been approximated to the one of the B^+ signal. Therefore, the results of the fit can be further improved optimising the shape modelling of these components.

7.2.2 Invariant Mass Fit in Different MVA Categories

The second strategy that has been investigated is the categorised MVA invariant mass fit. The MVA output is used to split the events into various categories (bins) with different signal over background ratios. The categories are fitted simultaneously to extract the parameters of interest using an un-binned extended maximum likelihood.

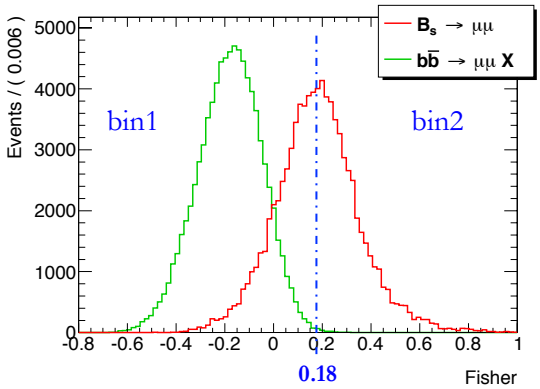


Fig. 7.12. *Fisher output distributions for background (solid green line) and signal (solid red line). The Fisher value of 0.18 divides the signal distribution into two bins with the same amount of signal events in each one.*

Tests have been performed using the Fisher trained with nine variables (section 7.1.1) and applying the additional selection to reduce the background. Two bins have been chosen (see figure 7.12) in order to have the same expected signal events in each one. Using the mass p.d.f.s for signal and background reported respectively in equation 7.6 and 7.7 we can build the likelihood for the fit:

$$\begin{aligned}\mathcal{L}_m &= \frac{N_s^{\text{bin 1}}}{N_s^{\text{bin 1}} + N_b^{\text{bin 1}}} \cdot \mathcal{M}_s^m(m \,|\, \mu_s^m, \sigma_{1,s}^m, \sigma_{2,s}^m, k_s^m) \\ &+ \frac{N_s^{\text{bin 2}}}{N_s^{\text{bin 2}} + N_b^{\text{bin 2}}} \cdot \mathcal{M}_s^m(m \,|\, \mu_s^m, \sigma_{1,s}^m, \sigma_{2,s}^m, k_s^m) \\ &+ \frac{N_b^{\text{bin 1}}}{N_s^{\text{bin 1}} + N_b^{\text{bin 1}}} \cdot \mathcal{M}_b^m(m \,|\, \alpha_b) + \frac{N_b^{\text{bin 2}}}{N_s^{\text{bin 2}} + N_b^{\text{bin 2}}} \cdot \mathcal{M}_b^m(m \,|\, \alpha_b)\end{aligned} \tag{7.12}$$

where the two double Gaussians of the B_s^0 signal mass distribution are fixed to the MC (figure 7.6 left), while the background shapes are left floating in the fit, but constrained to be the same in the two bins (the two related p.d.f.s have the slope parameter α_b in common). In this way the first bin, that is dominated by the continuum background, drives the shape in the second one in which the continuum is highly rejected by the MVA, and thus it does not have a well defined shape. Both the background and the signal normalisations are floating in each bin. An example of the fit procedure is reported in figure 7.13 that shows the results obtained with a single toy MC experiment generating 150 signal events in each bin and 100,000 total background events (that

correspond to 99,815 and 185 events respectively in the first and second bin, considering the selection on the Fisher output set at 0.18).

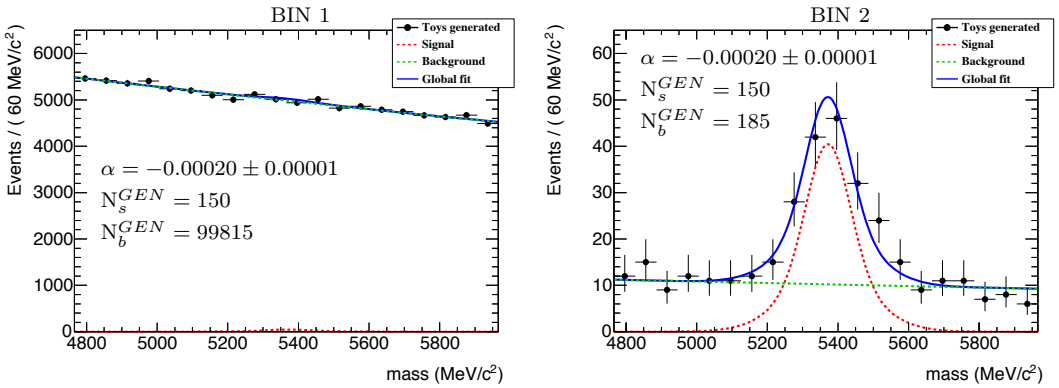


Fig. 7.13. *Example of categorised Fisher invariant mass fit obtained on one toy MC experiment generating 150 $B_s^0 \to \mu^+\mu^-$ signal events per bin, and 99,815 and 185 background events in the first and second bin respectively.*

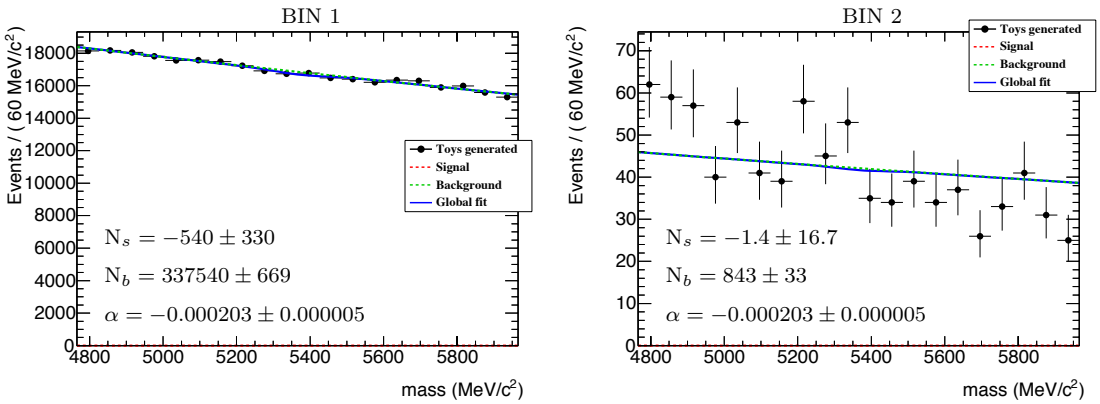
Parameter	Generated	From fit	Pull	
			Mean	Variance
N_s - bin 1	150	150 ± 182	-0.002 ± 0.006	1.009 ± 0.004
N_s - bin 2	150	150 ± 16	-0.026 ± 0.006	1.004 ± 0.004
N_{comb} - bin 1	99,815	99,815 ± 365	-0.0004 ± 0.0058	1.002 ± 0.004
N_{comb} - bin 2	185	185 ± 17	-0.037 ± 0.006	1.007 ± 0.004
α_{comb}	-0.000203	-0.000200 ± 0.000001	-0.006 ± 0.004	1.008 ± 0.003

Table 7.3. *Results of the toy MC experiments obtained for the fit in two Fisher categories generating 150 $B_s^0 \to \mu^+\mu^-$ signal events per bin, and 99,815 and 185 background events in the first and second bin respectively.*

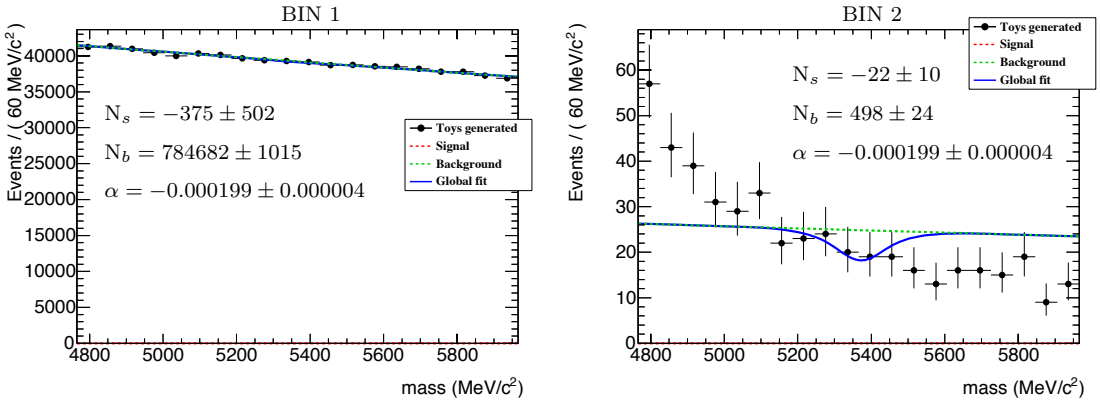
Toy MC experiments (30,000) have been produced with the same signal and background configuration reported above and used for the test shown in figure 7.13. The results obtained are summarised in table 7.3 (the related distributions of the yields and pulls are reported in appendix B.3.1). The pulls of the parameters of interest are compatible with zero and the biases on the signal event yield are less than 1 % proving the reliability of this fit procedure. Although the fit is able to extract the signal yield in the first bin, where the combinatorial background is dominant, the associated error is higher than 100 %. In these conditions, the first bin does not improve the signal sensitivity and it is used only to determine the background shape in the second bin.

7.2.3 Fit Tests on 2011 Un-blinded Data

After having checked the reliability of the two fit procedures using toy MC experiments, the next step is to test their performance using 2011 un-blinded data.



(a) *Categorised Fisher invariant mass fit*



(b) *Categorised BDT 2011 invariant mass fit*

Fig. 7.14. *Invariant mass fit, using 2011 un-blinded data, in two bins of the Fisher output trained with nine variables and applying the additional selection (**a**) and of the BDT 2011 output (**b**). The two bins have been chosen in order to have the same number of expected signal events in each one. The BDT classifier has a higher rejection power in the signal-like bin (BIN 2).*

The first series of tests was performed using the simultaneous fit in the MVA bins and the reason will be clear soon. Figure 7.14 shows the results obtained on 2011 un-blinded data with the Fisher trained with nine variables (applying the additional selection for the background reduction) and with the BDT used in the 2011 analysis (trained on the thirteen variables, without the additional selection).

Despite the bin definition for Fisher and BDT is different (related to their different performances), the background mass shape in the first MVA bin is the same in the two configurations due to also the very high statistic available in both cases, whereas in the second one the higher performances of the BDT are undeniable in the signal mass region. While in the second MVA bin the number of background events is more or less the same in the low-mass sideband (as it is clear from the first mass bins), in the signal region the higher rejection power of the BDT is evident. Since the second MVA bin is

highly signal-like, the effect of the additional selection is very mild (on the contrary to the first bin where the background is reduced by roughly a factor 2.5). What we learn from these plots, or better what is confirmed, is that in the most significant bin (in which the *signal-to-noise* ratio is higher) beside the continuum component, the lower mass region is dominated by a signal-like background. Since the first MVA bin is dominated by the continuum, the effect of these decays on its shape is negligible, but the same statement is not valid in the second bin. We conclude that, in order to be able to perform a reasonable fit, a second background component should be added to the model, to describe the exponential shape in the low mass region. The first MVA bin will still continue driving the shape of the combinatorial background in the second one.

Following also what has been reported in section 6.2.6, from now on, the background will be modelled using a linear function for the combinatorial component and an exponential function for the low-mass background.

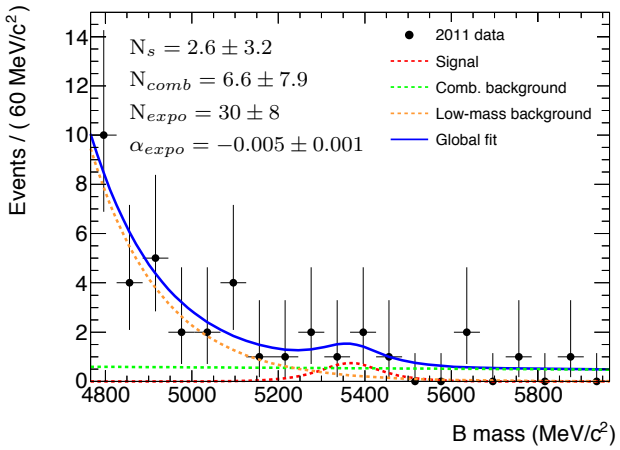


Fig. 7.15. *Invariant mass fit of 2011 un-blinded data for BDT > 0.118. The signal (red dashed line) is modelled using a double Gaussian fixed to the MC, the combinatorial background (green dashed line) is modelled with a linear function (fixed to the slope obtained from data without applying the BDT selection) and the low-mass background (orange dashed line) is modelled with an exponential free in the fit.*

At the time when these studies were performed, no 2012 MC to model neither the combinatorial nor the low-mass components was available. For that reason, a rough estimation of the shape of the low-mass background has been extracted from 2011 un-blinded data. Figure 7.15 reports the invariant mass fit of 2011 un-blinded data after having applied the optimised BDT selection (BDT > 0.118). The signal has been modelled using a double Gaussian with all parameters fixed to the MC, while the background is a combination of an exponential (floating in the fit) and a linear function (fixed to the slope obtained from data with no BDT selection applied). The value obtained for the exponential slope, 0.005 ± 0.001, will be used for the next tests and it is very close to the one that will be extracted using the final version of the fit

on 2012 data (see chapter 8).

7.2.4 Fit Tests on 2012-Like Analysis

In this section, the two strategies previously described are investigated in more detail using the two background components. In order to have a rough estimation of the expected significance, the number of background and signal events expected in 2012 has been used in the generation of the toy MC experiments. Since at the time when these studies were performed the final version of 2012 data and MC was not available, it was decided to continue relying on the 2011 datasets to model the shapes of signal and background, and the same selection efficiency obtained on the 4.9 fb^{-1} analysis has been used. While the first assumption is quite reasonable, beacuse we do not expect that the different reconstruction will produce big effects on the shape of the background and signal components, the second is only a rough approximation since the selection is one of the part of the analysis that will be better optimised with respect to the 2011 analysis[1]. However, with the intent of testing the fit procedures and give a coarse estimation of the expected significance, all these assumptions can be maintained for the time being.

The number of background events expected in the mass region [4766,5966] MeV has been extracted from a small sub-set of 2012 sideband data available, and then rescaled considering the total 2012 integrated luminosity. The number of signal events is estimated considering the SM BR for the $B_s^0 \to \mu^+\mu^-$ decay. Starting from the 2011 SES reported in table 5.7 and obtained applying the optimised BDT (BDT > 0.118) and signal mass window ($|\Delta m| < 121$ MeV) selection (see section 5.6), assuming a branching ratio of $(3.54 \pm 0.30) \cdot 10^{-8}$ [133], the number of expected events in 2011 is

$$\begin{aligned} N_s &= \frac{1}{\text{SES}} \cdot \text{BR}(B_s^0 \to \mu^+\mu^-) \\ &= \frac{1}{(2.07 \pm 0.26) \cdot 10^{-9}} \cdot (3.54 \pm 0.30) \cdot 10^{-9} = 1.7 \pm 0.3 \ . \end{aligned} \tag{7.13}$$

Considering the efficiency of the optimised BDT and mass window selection ($\sim 22\,\%$), in 2011 the number of expected signal events after only the baseline selection (section 5.4) is 7.7. This number can be rescaled considering the different luminosity and a 15 % increasing factor in the cross-section, giving a total number of expected events in 2012 equal to ~ 35 (34 events will be considered in the next tests in order to generate the same integer number in the two MVA bins). The normalisation of the low-mass component can be estimated from the fit of the un-blinded data reported in figure 7.15, obtained applying the optimised BDT selection for the 2011 analysis, that gives 30 low-mass background events. Assuming exactly the same behaviour of the signal in terms of MVA output, since the efficiency of the optimised BDT selection is $\sim 25\,\%$, the number of low-mass background events expected in 2011 is 120. Considering

[1] The selection adopted for the 2012 analysis is detailed in section 8.3. It benefits of a new method for the determination of the primary vertex associated to the B candidates and an improved muon selection that increases the $B_s^0 \to \mu^+\mu^-$ signal efficiency.

the same rescaling factor used for the signal, we expect around 550 events in 2012 data. Not having information neither about the MVA distribution nor about the mass shape modelling, in wanting to be quite conservative, roughly four times the low-mass background estimated above has been generated in the toy MC experiments.

7.2.4.1 2D Fit

The Fisher distributions for the signal and the combinatorial background have been modelled respectively with a double (equation 7.4) and a single (equation 7.5) Gaussian with the parameters fixed to the fit on the MC (reported in figure 7.5). Assuming that the low-mass background has a signal-like behaviour, the corresponding Fisher output can be modelled with the same function used for the B_s^0 signal.

The invariant mass of the $B_s^0 \to \mu^+\mu^-$ is described, as usual, using a double Gaussian (equation 7.6) fixed to the values obtained from the fit of the MC reported in figure 7.6 (left), while the combinatorial background is now modelled using a linear function

$$\mathcal{M}^m_{comb}(m|\alpha_{comb}) = \alpha_{comb} \cdot m \quad , \tag{7.14}$$

where the slope used in the generation of the toy MC experiments is extrapolated from 2011 un-blinded data (fit reported in figure 6.18 right) without the BDT selection applied. The low-mass background is modelled using an exponential

$$\mathcal{M}^m_{low-m}(m|\alpha_{low-m}) = e^{\alpha_{low-m} \cdot m} \quad , \tag{7.15}$$

where the slope generated is extracted from the fit reported in figure 7.15. The 2D model has now three contributions:

$$\begin{aligned}
\mathcal{M}^{F,m}_{s} &= \mathcal{M}^{\mathrm{F}}_{s}(F \,|\, \mu^F_s, \sigma^F_{1,s}, \sigma^F_{2,s}, k^F_s) \;\times\; \mathcal{M}^{m}_{s}(m \,|\, \mu^m_s, \sigma^m_{1,s}, \sigma^m_{2,s}, k^m_s) \\
\mathcal{M}^{F,m}_{comb} &= \mathcal{M}^{\mathrm{F}}_{comb}(F \,|\, \mu^F_{comb}, \sigma^F_{comb}) \;\times\; \mathcal{M}^{m}_{comb}(m \,|\, \alpha_{comb}) \\
\mathcal{M}^{F,m}_{low-m} &= \mathcal{M}^{\mathrm{F}}_{s}(F \,|\, \mu^F_s, \sigma^F_{1,s}, \sigma^F_{2,s}, k^F_s) \;\times\; \mathcal{M}^{m}_{low-m}(m \,|\, \alpha_{low-m})
\end{aligned} \tag{7.16}$$

and the 2-dimensional likelihood becomes

$$\begin{aligned}
\mathcal{L}^{2D}_{F,m} \;=\; & \frac{N_s}{N_s + N_{\mathrm{comb}} + N_{low-m}} \cdot \mathcal{M}^{F,m}_{s} \\
+ \; & \frac{N_{\mathrm{comb}}}{N_s + N_{\mathrm{comb}} + N_{low-m}} \cdot \mathcal{M}^{F,m}_{comb} + \frac{N_{\text{low-m}}}{N_s + N_{\mathrm{comb}} + N_{low-m}} \cdot \mathcal{M}^{F,m}_{low-m} \quad .
\end{aligned} \tag{7.17}$$

An extended un-binned maximum likelihood fit is then performed, fixing all Fisher parameters to the MC, and leaving floating all the normalisations as well as the background mass shape parameters. The performance of the 2D fit procedure have been tested running 30,000 toy MC experiments and the results are summarised in table 7.4 (the related plots are reported in appendix B.3.2.1). The pulls are all consistent with zero apart from the signal yield that shows a bias of 5.7 events ($\sim 17\,\%$). This big bias can in part be related to small number fluctuations that linearly decrease as the

statistic increases (e.g. the fit reported in section 7.2.1, performed on a higher statistic, does not present this intrinsic bias). On the other hand, part of the combinatorial and low-mass events can feed into the signal component confirmed by the negative bias on the respective yields.

Parameter	Generated	From fit	Pull	
			Mean	Variance
N_s	34	41.52 ± 40.5	0.14 ± 0.01	1.041 ± 0.004
N_{comb}	1,230,203	1,235,067 ± 1120	-0.024 ± 0.001	1.011 ± 0.004
N_{Low-m}	2,200	2,200 ± 96	-0.012 ± 0.006	1.013 ± 0.004
α_{comb}	-0.121	-0.122 ± 0.002	-0.001 ± 0.006	1.019 ± 0.004
α_{low-m}	-0.005	-0.0050 ± 0.0001	-0.049 ± 0.001	1.055 ± 0.005

Table 7.4. *Results of the toy MC experiments for the 2D fit, obtained simulating the statistic expected in 2012.*

7.2.4.2 Categorised MVA Mass Fit

Tests have been performed using the Fisher trained with nine variables (section 7.1.1) applying the additional selection to reduce the background. Two bins have been chosen as reported in section 7.2.2. The signal p.d.f is modelled with a double Gaussian (equation 7.6), the background by a linear function (equation 7.14), while the low-mass decays by an exponential (equation 7.15). The global likelihood is therefore

$$
\begin{aligned}
\mathcal{L}_m \;=\; & \frac{N_s^{\text{bin 1}}}{N_s^{\text{bin 1}} + N_{comb}^{\text{bin 1}} + N_{low-m}^{\text{bin 1}}} \cdot \mathcal{M}_s^m(m \,|\, \mu_s^m, \sigma_{1,s}^m, \sigma_{2,s}^m, k_s^m) \\
+ \; & \frac{N_s^{\text{bin 2}}}{N_s^{\text{bin 2}} + N_{comb}^{\text{bin 2}} + N_{low-m}^{\text{bin 2}}} \cdot \mathcal{M}_s^m(m \,|\, \mu_s^m, \sigma_{1,s}^m, \sigma_{2,s}^m, k_s^m) \\
+ \; & \frac{N_{comb}^{\text{bin 1}}}{N_s^{\text{bin 1}} + N_{comb}^{\text{bin 1}} + N_{low-m}^{\text{bin 1}}} \cdot \mathcal{M}_{comb}^m(m \,|\, \alpha_{comb}) \\
+ \; & \frac{N_{comb}^{\text{bin 2}}}{N_s^{\text{bin 2}} + N_{comb}^{\text{bin 2}} + N_{low-m}^{\text{bin 2}}} \cdot \mathcal{M}_{comb}^m(m \,|\, \alpha_{comb}) \\
+ \; & \frac{N_{low-m}^{\text{bin 1}}}{N_s^{\text{bin 1}} + N_{comb}^{\text{bin 1}} + N_{low-m}^{\text{bin 1}}} \cdot \mathcal{M}_{low-m}^m(m \,|\, \alpha_{low-m}) \\
+ \; & \frac{N_{low-m}^{\text{bin 2}}}{N_s^{\text{bin 2}} + N_{comb}^{\text{bin 2}} + N_{low-m}^{\text{bin 2}}} \cdot \mathcal{M}_{low-m}^m(m \,|\, \alpha_{low-m}) \quad ,
\end{aligned}
\tag{7.18}
$$

where the two double Gaussians of the B_s^0 signal mass shape are fixed to the MC (figure 7.6 left), while the combinatorial and low-mass background shapes are floating in the fit, but constrained to be the same in the two bins (the parameters α_{comb} and α_{low-m} are in common in the related p.d.f.s of the two bins). The mass shape of the first bin

drives the combinatorial background, while the one of the low-mass background is constrained by the second bin. The normalisations of the two backgrounds and the signal are floating in each bin. In the toy MC experiments 17 signal and 1100 low-mass events per bin have been generated, while the total background expected in 2012 is split into 1,227,881 events in the first bin and 2,322 in the second one.

Parameter	Generated	From fit	Pull	
			Mean	Variance
N_s - bin 1	17	28 ± 724	0.018 ± 0.013	1.013 ± 0.009
N_s - bin 2	17	17.7 ± 31.5	0.011 ± 0.013	1.021 ± 0.009
N_{comb} - bin 1	1,227,881	1,227,829 ± 2373	-0.010 ± 0.013	1.005 ± 0.009
N_{comb} - bin 2	2,322	2319 ± 99	-0.011 ± 0.014	1.08 ± 0.01
N_{low-m} - bin 1	1100	1133 ± 1626	-0.019 ± 0.013	0.987 ± 0.009
N_{low-m} - bin 2	1100	1104 ± 88	-0.027 ± 0.014	1.06 ± 0.01
α_{comb}	-0.121	-0.121 ± 0.003	0.009 ± 0.013	1.007 ± 0.009
α_{low-m}	-0.005	-0.0051 ± 0.0005	-0.024 ± 0.013	1.039 ± 0.010

Table 7.5. *Results of the toy MC experiments obtained for the fit in two Fisher categories simulating 2012 statistic.*

The results obtained generating 30,000 toy MC experiments are summarised in table 7.5, while the distributions of the parameters of interest and the related pulls are reported in appendix B.3.2.2.

The signal bias in the second bin is about 2 % and the pulls of the other parameters are compatible with zero, showing the reliability of the fit in the most significant bin. As already shown in section 7.2.2 for the model with only one background, the very small *signal-to-noise* ratio in the first bin leads to an error on the signal (and low-mass) yield that is higher than the corresponding mean value. A test has been performed fixing the signal and the low-mass normalisations in the first bin to the related expected values to quantify possible biases, especially on the shape of the low-mass background. The results, reported in appendix B.3.2.2, show how the effects on the fit parameters and yields are negligible.

7.2.4.3 Fit Significance

In order to assess and compare the performances of the fit strategies analysed in the previous sections, the significance, defined as the likelihood ratio between the *signal+background* and the *background-only* hypothesis,

$$Z = \sqrt{-2 \cdot \log\left(\frac{\mathcal{L}(0, \hat{\hat{\theta}}(0))}{\mathcal{L}(\hat{\mu}, \hat{\theta})}\right)} = \sqrt{-2 \cdot \log\left(\lambda(0)\right)} \quad , \tag{7.19}$$

was studied. In the denominator the maximum likelihood estimators $\hat{\mu}$ and $\hat{\theta}$ are obtained maximising the likelihood with respect to both the number of signal events

μ and the fit parameters θ, while in the numerator the estimator $\hat{\hat{\theta}}(0)$ is obtained maximising the likelihood fixing μ to zero.

Figure 7.16 reports the significance for the two different configurations described in the previous sections considering the number of signal and background events expected in 2012. The results show how the performances of the two typologies of fit are very similar.

In addition, in order to quantify how much the performance of the MVA classifier may affect the $B_s^0 \to \mu^+\mu^-$ sensitivity, another test has been performed simulating an increased separation between signal and background events, obtained shifting the mean of the signal classifier distribution towards higher values. Table 7.6 summarises the separations obtained (evaluated using equation 7.3) and the related number of signal and background events expected.

Shift	**Median signal distribution**	**Separation**	$\mathbf{N}_s$		$\mathbf{N}_{comb}$	
			bin 1	**bin 2**	**bin 1**	**bin 2**
0.0	0.175	0.685	17	17	1227881	2322
0.073	0.258	0.795	17	17	1229875	328
0.100	0.275	0.829	17	17	1229961	242

Table 7.6. *Separation between signal and background distributions of the Fisher output simulating a shift of the mean value of the signal classifier. The related expected number of background events is also reported. A separation of 0.795 corresponds to the one obtained with the BDT trained for the 2011 analysis.*

Fit Strategy	**Shift**	**Separation**	**Significance**
2D	0.0	0.685	1.14 ± 0.73
2D	0.073	0.795	1.32 ± 0.81
2D	0.100	0.829	1.38 ± 0.83
Categorised MVA	0.0	0.685	0.97 ± 0.66
Categorised MVA	0.073	0.795	1.22 ± 0.78
Categorised MVA	0.100	0.829	1.28 ± 0.80

Table 7.7. *Signal significance expected for the two fit strategies. A separation of 0.795 corresponds to the one obtained with the BDT trained for the 2011 analysis.*

Table 7.7 summarises the signal significance for the two fit strategies obtained with the default Fisher and simulating an increased separation. These results show how the significance, with the current fit configurations, is mildly affected by an improved background rejection and they confirm the necessity of an optimisation of the procedures to obtain a competitive sensitivity in the $B_s^0 \to \mu^+\mu^-$ channel. In section 7.2.4.5 an overview of possible improvements will be given.

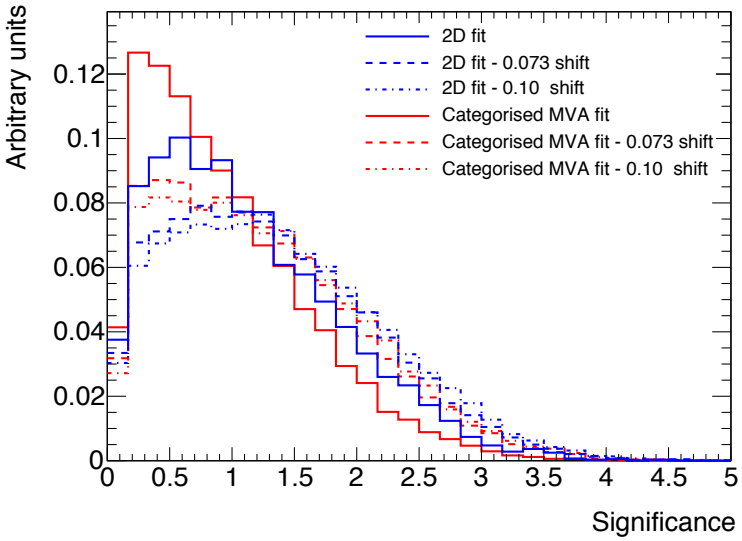


Fig. 7.16. *Signal significance for the 2D fit (blue lines) and the categorised MVA fit (red lines). The solid lines correspond to the results obtained considering the Fisher output trained with nine variables (applying the additional selection for the background reduction), while the dashed and dashed-dotted lines show the significance increasing the separation between signal and background.*

7.2.4.4 ATLAS/CMS Performance Comparison

In order to try to understand the low fit sensitivity shown in the previous section, a comparison with the CMS and LHCb results available was made. Although differences in the detector design, CMS is the experiment of the two that most resembles ATLAS. Therefore, the idea is to extract the number of background and signal events from the CMS mass plot shown in the 2012 published analysis [7], generate mock data from these numbers, perform the invariant mass fit and finally check the effect of the different ATLAS trigger efficiency and mass resolution on the signal sensitivity.

CMS analysis is based on the simultaneous fit of twelve different BDT bins, subdivided in three $|\eta|$ regions to take into account of the different di-muon mass resolution in the barrel and in the end-caps. The twelve BDT categories are fitted simultaneously to extract the total number of signal events. The fit method is not detailed in the public documentation, but the methodology is very similar to the one we were developing in ATLAS.

The mass resolution of CMS is around 40 MeV in the barrel region ($|\eta| < 1.4$) and 74 MeV in the end-caps, compared to the one of ATLAS that is approximately 60/78/107 MeV in the three $|\eta|$ regions [0, 1.0]/[1.0, 1.5]/[1.5, 2.5] (see section 8.4 for more details). In addition, the CMS trigger is much more efficient with respect to ATLAS, 88 % versus 44 % respectively, in part also related to a different L1 trigger. For 8 TeV data, in the CMS barrel region, the p_T thresholds were loosened to 3 GeV (requiring $p_T^{\mu\mu} > 4.9$ GeV), while at the high level trigger two muons, each with $p_T > 4$

GeV ($p_T^{\mu\mu} > 7$ GeV) and $|\eta| < 2.2$, were selected. In October 2011 (after the LHC technical stop), in order to keep the rate under control despite the increased luminosity, ATLAS raised the thresholds on the L1 muons from 0 to 4 GeV (see section 5.3), maintaining essentially the same selection at the HLT. The lower L1 thresholds adopted by CMS allow to improve the efficiency in selecting low-p_T muons effectively increasing the available signal statistic in the $B_s^0 \to \mu^+\mu^-$ channel.

Component	**Bin 1**	**Bin 2**	**Bin 3**	**Bin 4**
N_{B_s}	7	7	7	7
N_{B^0}	1	1	1	1
$N_{B\to hh'}$	1	1	1	1
N_{comb}	4800	400	28	2
$N_{B_{low-m}}$	40	32	25	10

Table 7.8. *Number of signal and background events extracted from the CMS mass plot [7].*

The number of events for signal and background measured by CMS has been roughly estimated counting the events in the di-muon mass plots contained in the published analysis performed on 2012 data [7]. For these tests, only the four bins in the barrel have been considered. The normalisations for the signal and the background contributions, extracted from the CMS mass plots in the four BDT bins, are summarised in table 7.8 and they have been used to generate the toy MC experiments to estimate the expected significance. The continuum has been modelled with a linear function, the low-mass background with an exponential and the signal (both B_s^0 and B^0) with a single Gaussian. The peaking background $B \to hh'$ component has been generated using a single Gaussian, but its contribution is not included in the fit likelihood for sake of simplicity[2]. All the shapes are extracted from ATLAS data and MC. In the fit, the signal B_s^0 and B^0 shapes are fixed to the MC, while the background slopes are free to float.

Figure 7.17 shows the significance obtained considering the CMS (40 MeV) or ATLAS (60 MeV) di-muon mass resolution in the barrel region. The significance obtained from the toy MC experiments (4.2 ± 1.1) considering the CMS mass resolution is compatible with the one published and drops to a value of 2.8 ± 1.0 using exactly the same fit configuration and the same number of events, but increasing the mass resolution to the one expected in the ATLAS barrel.

In a second test, only the number of generated events is changed in order to simulate the effect of the different trigger efficiencies. The numbers seeding the toy MC experiments are half of the ones reported in table 7.8. Figure 7.18 shows the results obtained: the lower efficiency drops the significance to 2.4 ± 1.0 (considering a mass resolution of 40 MeV) that in turn it is further reduced to 1.6 ± 0.9 considering the ATLAS mass resolution of 60 MeV.

[2]The bias on the signal yield introduced by this assumption is very small and acceptable for the purposes of these tests.

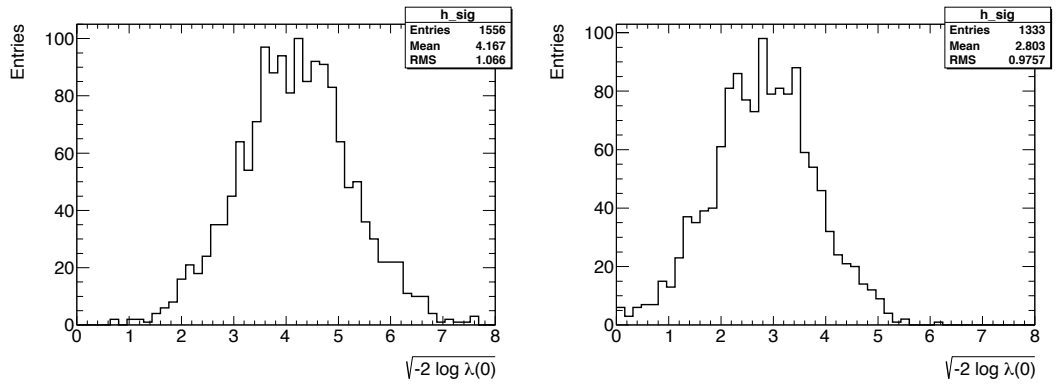


Fig. 7.17. *Fit significances obtained considering the CMS (left) and ATLAS (right) di-muon mass resolution, generating the number of events extracted from the CMS mass plot [7].*

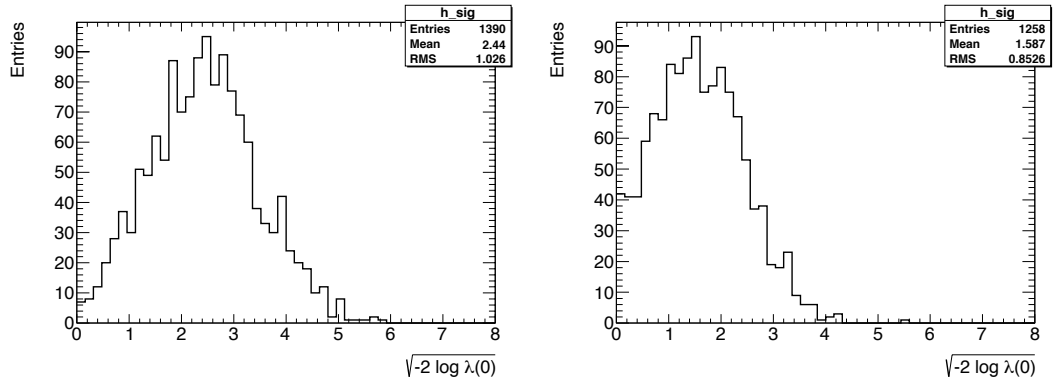


Fig. 7.18. *Fit significances obtained considering the CMS (left) and ATLAS (right) di-muon mass resolution, generating half of the statistic extracted from the CMS mass plot [7].*

A summary of the trigger efficiency and mass resolution effects is reported in table 7.9 that shows how the CMS sensitivity is degraded of about $2.6\,\sigma$ considering the performance of the ATLAS detector. Although these tests are quite rough, they allow to better understand our weaknesses. Unfortunately nothing can be done on data to recover the lost sensitivity, but it will be shown how an optimisation of the mass fit and a precise knowledge of the background contributions allow to partially fill the gap between the two performances.

7.2.4.5 Possible Optimisation of the Fits Performance

The two fit strategies analysed in the previous sections showed to be quite stable and reliable in extracting the signal yield, but they have to be optimised to maximise the $B_s \to \mu^+\mu^-$ significance.

To be able to perform a 2D fit we need to use the whole range for the MVA variable. Despite the good performance shown by this method, the *signal-to-noise* ratio is still quite small to obtain a high sensitivity, and for that reason the background has to be

Mass resolution \ Trigger efficiency	**88 %**	**44 %**
40 MeV	4.2 ± 1.1	2.4 ± 1.0
60 MeV	2.8 ± 1.0	1.6 ± 0.9

Table 7.9. *Summary of the effects of the ATLAS trigger and mass resolution on the expected significance obtained simulating the CMS statistic.*

highly reduced. For this purpose, two different strategies can be considered.

The first one is to introduce, before any MVA selection, a selection based on some "event-shape" discriminating variables. One fundamental difference between B meson signal and combinatorial background is the kinematic of their underlying production at the LHC. After a baseline kinematic selection, additional background rejection could be possible by exploiting differences in the angular distributions of the muons produced in signal and background events being the latter, as we have already seen in section 5.2, composed of processes $b\bar{b}(b\bar{b}b\bar{b}, b\bar{b}c\bar{c}) \rightarrow \mu^+\mu^- X$ with muons produced mainly from semi-leptonic b and c decays.

To perform this background rejection, different variables can be used. Given a collection of N momenta, one can define the unit vector T, called *thrust*, along which their total projection is maximal

$$T = \frac{\sum_{i=1}^{N} |\boldsymbol{T} \cdot \boldsymbol{p}_i|}{\sum_{i=1}^{N} |\boldsymbol{p}_i|} \quad . \tag{7.20}$$

This vector can then be used to define various related variables, like $|\cos\theta_T|$, where θ_T is the angle between the thrust axis of the momenta of the B candidate decay particles and the thrust axis of all the other particles in the event that are not associated to the B. When the decay particles are distributed isotropically, the thrust axis is randomly distributed and therefore the $|\cos\theta_T|$ has a uniform distribution between 0 and 1. On the contrary, in case of strongly directional events the thrusts are strongly collimated leading to a distribution that is peaked at large values. Another variable with a high discrimination power is the *sphericity* tensor defined as

$$S^{\alpha,\beta} = \frac{\sum_{i=1}^{N} p_i^{\alpha}\, p_i^{\beta}}{\sum_{i=1}^{N} |\boldsymbol{p_i}|^2} \quad , \tag{7.21}$$

(where $\alpha, \beta = x, y, z$) that provides a three-dimensional representation of the spatial distribution of $\boldsymbol{p_i}$. In case of an isotropic distribution the three eigenvalues λ_k have the same magnitude, whereas for a planar distribution one of the eigenvalues is significantly smaller with its eigenvector orthogonal to that plane. When the distribution is very directional, the eigenvector that is oriented in that direction has an eigenvalue that is much larger than the other two. From the above definition we can derive the scalar

sphericity

$$S = \frac{3}{2}\left(\lambda_2 + \lambda_3\right) \quad , \tag{7.22}$$

where λ_2 and λ_3 are the lowest eigenvalues. When the momenta are isotropically distributed S assumes values close to 1, while if the distributions are very collimated then $S = 0$. Another useful set of variables are the *Fox-Wolfram moments* [138] defined as

$$H_k = \sum_{i\,,j}^{N} |\boldsymbol{p}_i|\,|\boldsymbol{p}_j|\,P_k(\cos\theta_{ij}) \quad , \tag{7.23}$$

where θ_{ij} describes the angle between the momenta $\boldsymbol{p}_i$ and $\boldsymbol{p}_j$, while P_k is the k-th order Legendre polynomial. For strongly collimated events, H_k assumes values close to zero (one) for odd (even) values of k.

The second strategy would be to optimise the additional selection described in section 6.1.3. One may try to train a simple MVA classifier using the same discriminating variables (or a smaller subset) adopted for the final selection and produce a subsample optimising a BDT selection that retains at least 90/95 % of the signal events. The final MVA studies will be then performed on this dataset. In this way, ideally, it would be possible to use the full MVA distribution and improve the sensitivity, being the number of background events highly reduced by the first MVA selection.

Unfortunately, due to a tight analysis schedule, both options could not be tested in detail. In particular, the first one requires to have access to all the information from the tracks produced in an event, that were not saved in the default ntuples used by the group. When the hypothesis was considered the production of the final version of the datasets for the 2012 analysis was already started, including the new background MC that requires very long processing times. Redoing the production was not a feasible option. Nonetheless, these strategies constitute an interesting possible optimisation of the selection that could be investigated for Run 2 data.

Due to the impossibility of consistently reducing the current levels of background, the 2D fit was not consider in the final strategy for the signal yield extraction in 2012 analysis, while the efforts were concentrated in the optimisation of the fit in the MVA bins.

All the tests reported in the previous sections showed how the whole MVA spectrum, in the current conditions, can not be used neither for the 2-dimensional fit nor for the fit in the MVA bins (unless a strong background rejection, as the one reported above, is introduced). For the latter strategy the idea is to reduce the width of the bins considering only the most signal-like part of the classifier in which the *signal-to-noise* ratio assumes reasonable values. This will not allow to use the whole statistic, but the reasonable loss of signal events ($< 30\,\%$) is compensated by a significant increase of the signal sensitivity. In addition, having more than one background contribution, one may consider the hypothesis of using more than two bins each of them providing specific information on the different components. This is the strategy adopted for the 2012 analysis and it will be discussed in detail in chapter 8.

In order to reduce the error on the signal yield in the bins in which the *signal-to-noise* ratio is lower, instead of evaluating the $B_s^0 \to \mu^+\mu^-$ normalisation separately in each bin, it is possible to extract a global normalisation factor fixing the fraction of signal events in each bins. The same strategy can be adopted for the background normalisation. For the 2012 analysis three bins in the BDT output (with the same expected number of signal events in each one) have been chosen for the final version of the invariant mass fit, but more details will be provided in section 8.10.

What is important to stress is that the simultaneous fit is a powerful tool to increase the significance of the analysis, but several meticulous checks have to be done to validate the whole procedure. For example, performing the simultaneous fit of the continuum across the bins we are constraining the shape of the background to be the same in each MVA category, and therefore we are implicitly assuming that there is no correlation (or the correlation is rather small) between the invariant mass and the MVA output. Neglecting these effects can lead to a wrong estimation of the background. Also the assumption of having the same MVA efficiency in each bin must be carefully checked.

Chapter 8

$B_s \to \mu^+\mu^-$ Analysis on 2012 Data

This chapter describes the $B_s^0 \to \mu^+\mu^-$ 2012 analysis performed on the full Run 1 dataset. The aim is to obtain the best ATLAS result on the $B_s^0 \to \mu^+\mu^-$ final state and a first ATLAS limit on the $B^0 \to \mu^+\mu^-$ channel.

In the next sections, I will give an overview of the whole analysis, showing the various improvements that have been introduced, but I will focus on the trigger and background modelling as well as on the fit for the extraction of the $B_s^0 \to \mu^+\mu^-$ signal yield.

In particular I will describe the generation of the semi-leptonic samples (section 8.2.1) for which I performed detailed studies and I developed a signal filter (now integrated in the ATLAS software framework) that allows to select the interesting events, optimising the generation. All these studies proved to be particularly useful to optimise also the other MC samples, for which the cuts at generator level have been carefully tuned. I checked the data-MC agreements for the discriminating variables used in the 2012 BDT (section 8.7) and I introduced a pile-up and kinematical re-weighting to correct for the inaccuracies present in the background MC (section 8.2.2.1). All the studies I performed for the optimisation of the $B_{s,d} \to \mu^+\mu^-$ invariant mass fit will be extensively detailed in section 8.10 and a preliminary result obtained with the un-blinded 2011 and 2012 data will be discussed.

8.1 Introduction

The strategy for this updated analysis is revisited with respect to the previous versions, focusing on the possibility of a measurement of the $B_s^0 \to \mu^+\mu^-$ branching fraction. As detailed in section 1.6, the newest theoretical prediction on these branching ratios are $\mathrm{BR}(B_s^0 \to \mu^+\mu^-) = (3.65 \pm 0.23) \cdot 10^{-9}$ and $\mathrm{BR}(B_d^0 \to \mu^+\mu^-) = (1.06 \pm 0.09) \cdot 10^{-10}$ [3]. CMS and LHCb now have a joint $6.2\,\sigma$ measurement for the $B_s^0 \to \mu^+\mu^-$ final state giving an average branching ratio of $(2.8 \pm 0.7) \cdot 10^{-9}$ [28]. As already seen in the previous chapter, ATLAS has limited trigger efficiency and mass resolution,

resulting in a degraded sensitivity to these decays. In order to exploit at best ATLAS data, the analysis has been updated in several aspects benefiting from the improved statistics, taking full advantage of the very large Monte Carlo samples and re-optimising the background rejection. The peaking background will be a major contributor under the signal peaks, so extra effort has been invested in reducing the muon fake rate introducing a specific selection based on a BDT classifier that allows to keep the level of fake muons under 1 ‰ (see section 8.6).

However, the main idea guiding the new analysis is to increase the sensitivity to the signal by using a mass fit on the widest possible set of events (see section 8.10). Thus a loose selection will be applied to retain a maximum number of signal events, and the final fit will distinguish between the signal and the various backgrounds. Another novelty of this analysis is the explicit study of partially reconstructed B decays as they contribute to the background in the low-mass sideband.

The master formula for the extraction of the branching ratio is similar to the one detailed in section 5.1, but it is amended to take into account the use of different triggers:

$$\begin{aligned} \mathrm{BR}\left(B_s^0 \to \mu^+\mu^-\right) = \mathrm{BR}\left(B^\pm \to J/\psi K^\pm \to \mu^+\mu^- K^\pm\right) \cdot \tfrac{f_u}{f_s} \cdot N_{\mu^+\mu^-} \cdot \\ \cdot \left(\textstyle\sum_k N^k_{J/\psi K^\pm}\, \alpha_k\, \frac{(A\times\varepsilon)^k_{\mu^+\mu^-}}{(A\times\varepsilon)^k_{J/\psi K^\pm}} \right)^{-1}, \end{aligned} \tag{8.1}$$

where the index k runs on the trigger categories used in the analysis (see section 8.5). The α_k parameter takes into account the trigger prescaling and the luminosity factors applied to all the events.

The B^+ yield for the reference channel has been extracted with an un-binned maximum likelihood fit to the invariant mass distribution. The fit is performed simultaneously on data and MC and will be described in section 8.8.

A similar procedure used in the previous analysis [106] has been adopted for the estimation of the acceptance times efficiency ratio in 2012 analysis that is now evaluated for each trigger category (see section 8.9).

8.2 Data and MC Samples

The 2012 analysis uses $\sqrt{s} = 8$ TeV and $\sqrt{s} = 7$ TeV pp collision data collected by ATLAS with stable LHC beams during 2011 and 2012 data taking period (more details about the combination of the two datasets are reported in section 8.10.9). Specific data-quality requirements have been applied in order to retain only the events in which both the tracking and muon detectors were fully operational, which is essential for the reconstruction of B_s^0 and $B^\pm$ mesons. This leads to a total integrated luminosity for the combination of 2011 and 2012 datasets of about 25 fb^{-1}.

The MC samples adopted for the 2012 analysis are listed in table 8.1, produced with the 2012 MC tuning and detector simulation. The novelty with respect to the 2011 analysis is the introduction of three exclusive samples to model the semi-leptonic decays

in the low-mass region and a high statistic MC sample to describe the combinatorial background ("four-corners"), that will be discussed in more details in sections 8.2.1 and 8.2.2 respectively. In the MC generation, the modelling of the proton collisions, hard processes, hadronization and the consequently decay of the particles is performed by PYTHIA8B [144] except for the reference channel, the semi-leptonic and the "four-corners" MCs where the decay is performed by EVTGEN [140] (for more detail about the generation see section 8.2.1.1). Only the $B_s^0 \to \mu^+\mu^-$ signal and the two samples used for the fake muons studies ($B \to hh'$ and $\Lambda_b^0 \to ph$) are produced using the full detector simulation GEANT4 [116], while AtlfastII [117] was used for all the other MCs (see section 5.3 for more details about the fast detector simulation).

MC sample	Generation	Simulation	Events
$B_s^0 \to \mu^+\mu^-$	PYTHIA8B	GEANT4	$\sim$ 400 k
$B^0 \to \mu^+\mu^-$	PYTHIA8B	Atlfast II	$\sim$ 400 k
$B^+ \to J/\psi\,(\mu^+\mu^-)\,K^+$	PYTHIA8B + EVTGEN	Atlfast II	$\sim$ 2.5 M
$B^- \to J/\psi\,(\mu^+\mu^-)\,K^-$	PYTHIA8B + EVTGEN	Atlfast II	$\sim$ 2.5 M
$B^+ \to J/\psi\,(\mu^+\mu^-)\,\pi^+$	PYTHIA8B + EVTGEN	Atlfast II	$\sim$ 1 M
$B_s^0 \to J/\psi\,(\mu^+\mu^-)\,\phi\,(K^+K^-)$	PYTHIA8B	Atlfast II	$\sim$ 40 M
$B \to hh'$	PYTHIA8B	GEANT4	$\sim$ 4 M
$\Lambda_b^0 \to p\,h$	PYTHIA8B	GEANT4	$\sim$ 2 M
"Four-Corners"	PYTHIA8B + EVTGEN	Atlfast II	$\sim$ 1.3 G
$B^0 \to \pi\,\mu\,\nu$	PYTHIA8B + EVTGEN	Atlfast II	$\sim$ 500 k
$B_s \to K\,\mu\,\nu$	PYTHIA8B + EVTGEN	Atlfast II	$\sim$ 100 k
$\Lambda_b^0 \to p\,\mu\,\nu$	PYTHIA8B + EVTGEN	Atlfast II	$\sim$ 100 k
$b\bar{b} \to J/\psi\,(\mu^+\mu^-)X$	PYTHIA8B	Atlfast II	$\sim$ 1 M

Table 8.1. *MC samples produced for the 2012 analysis.*

The MC samples have been reweighed, using the GL corrections and the DD weights, to be corrected for small effects related to the generation level selection and to correct the kinematic of the B candidates. The weights have been evaluated with the same procedure adopted in the previous analysis [106] (see sections 5.3 and 5.5.4).

8.2.1 The Semi-Leptonic Decays Background Sample

In chapter 7 it was shown how the low-mass sideband is supposed to be populated by semi-leptonic decays with a signal-like behaviour. Despite the very low fraction for $h \to \mu$ misidentification, this kind of background was still entering in the selected

events in LHCb and CMS analyses [5, 6], therefore explicit tests have been performed in ATLAS.

Since the MC for these samples were never produced before in ATLAS, specific studies were dedicated to the generation of three different modes that are expected to be the most significant ones.

- $B^0 \to \pi\,\mu\,\nu$, with BR $\sim 1.4 \cdot 10^{-4}$
- $B_s^0 \to K\,\mu\,\nu$, BR not measured, but assumed to be $\sim 10^{-4}$
- $\Lambda_b^0 \to p\,\mu\,\nu$, BR not measured, but assumed to be $\sim 10^{-4}$

8.2.1.1 The Event Generation

For LHC experiments the events simulation requires the modelling of the initial *pp* collisions, the production of quarks, gluons and gauge bosons from the collisions and the subsequent hadronization and thus the decay of these hadrons into particles that can be measured by the detectors. The output of these generations has the form of a list of particles, their related momentum and energy, their parent particles and the daughters into which they decay, and the position of the production and the decay vertices.

The LHC experiments have a wide range of event generators depending on the particular kind of physics under study. For B-physics purposes, the core package is PYTHIA[1] [114, 139] that can handle the simulation of the initial partonic collision described by perturbative QCD. There are also a range of specialised packages and among all of them EVTGEN [140] was specifically created for B-meson decays. EVTGEN is not capable of modelling the proton collisions, hard processes or hadronization that have still to be carried out with PYTHIA, but has a number of unique features which make it particularly suitable for the simulation of B-decays. Decay amplitudes, instead of probabilities, are used for the simulation of the decays. The amplitude for each node in a decay tree is used to simulate the entire decay chain, including all angular and time-dependent correlations.

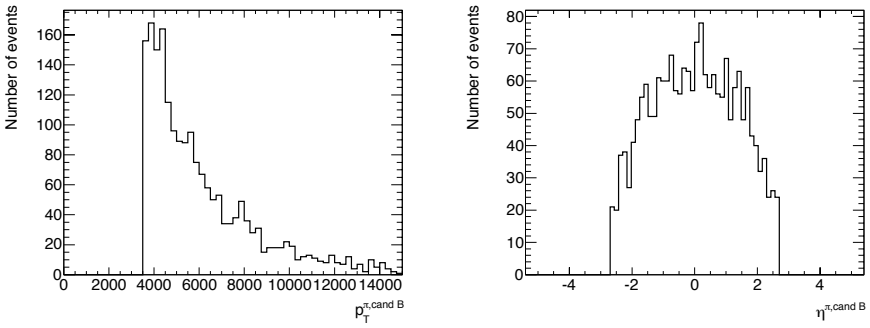
The results of the event generation are then written in a common format, known as HepMC [141]. Additional algorithms, known as *Generator Filters* may be used to select the events passing the user requirements and are also useful to simulate the action of the trigger systems.

8.2.1.2 The BSignalFilter Algorithm

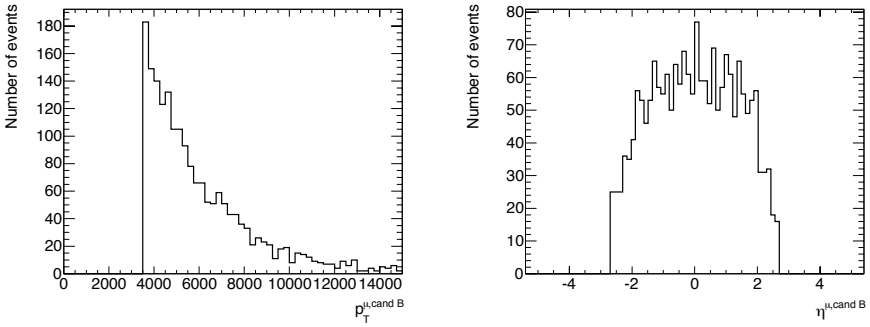
The *BSignalFilter* is an algorithm that has been implemented in the ATLAS software framework Athena [142, 143] and that has essentially two roles. The first one is to remove from the events generated by PYTHIA all those containing un-decayed

[1] Pythia, known as Oracle of Delphi, was the name given to any of the priestesses of the Temple of Apollo in Delphi (on the mount Parnassus). The purpose of Pythia was to share a glimpse of the future, offering several prophecies, all of them vague, misleading, and in some cases true (it is said she predicted the Trojan war).

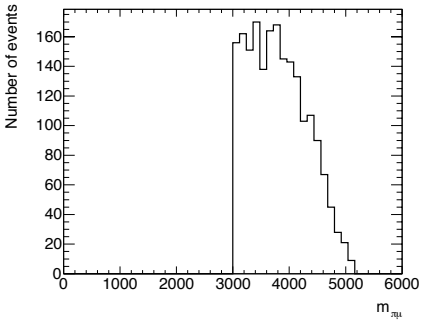
unstable particles (the presence of these particles indicates that all the possible related decay modes have been switched off by the user). In this case, the event is removed before the relative information is saved. The second role is to simulate the presence of a trigger and therefore remove particles kinematically not relevant (e.g. applying a selection in η and/or p_T). In addition, a selection on the hadron-muon invariant mass has been implemented in the algorithm in order to retain only the events that are generated with a mass of the final state hadron-muon couple within a certain range defined by the user.



(a) *Pion p_T (left) and η (right).*



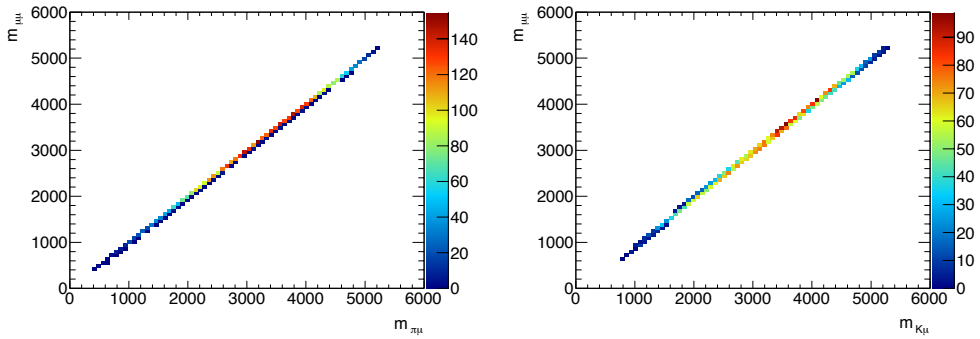
(b) *Muon p_T (left) and η (right).*



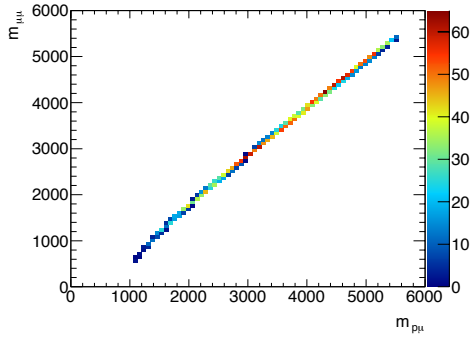
(c) *Invariant mass of the π-μ system.*

Fig. 8.1. *Test sample of 2000 $B^0 \to \pi\,\mu\,\nu$ generated events selected using the BSignalFilter, requiring a pion and a muon with $p_T > 3.5$ GeV and $|\eta| < 2.7$, as well as an invariant mass selection on the hadron-muon pair higher than 3.0 GeV.*

Before integrating the *BSignalFilter* in the Athena framework, its reliability and performances have been checked. Samples have been generated using PYTHIA8B [144] and then decaying the particles with EVTGEN. Figure 8.1 shows the results obtained generating a small sample (2000 events) of $B^0 \to \pi\,\mu\,\nu$ and simulating a trigger with $|\eta|$ lower than 2.7 and with a p_T threshold at 3.5 GeV, as well as an invariant mass selection on the pion-muon system higher than 3.0 GeV. The results show how the cuts are correctly applied to the generated events.



(a) $B^0 \to \pi\,\mu\,\nu$ **(b)** $B_s^0 \to K\mu\,\nu$

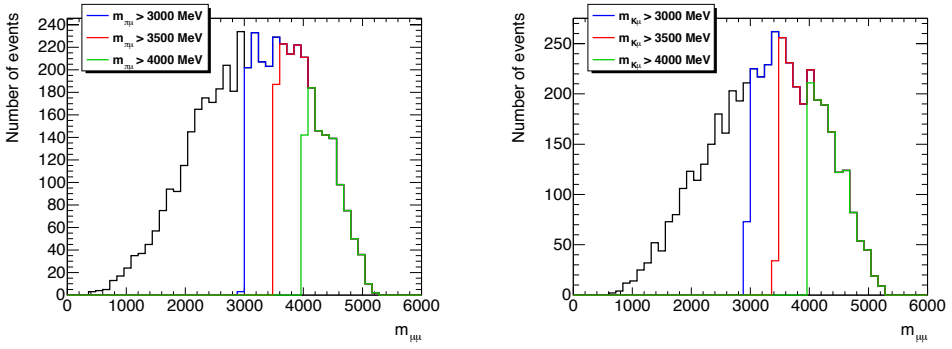


(c) $\Lambda_b^0 \to p\,\mu\,\nu$

Fig. 8.2. *Scatter plot of the di-muon versus hadron-muon system mass at generator level for each of the three semi-leptonic channels.*

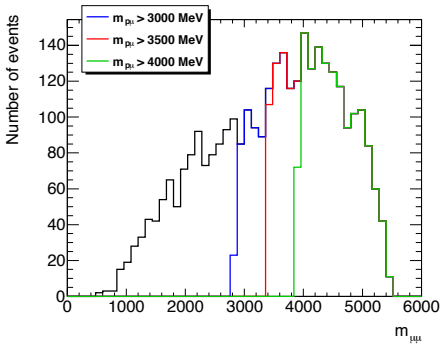
Since at reconstruction level the calculation of the muon mass is associated to the misidentified hadron, to establish the correct invariant mass range to be applied to the MC samples in order not to bias the distributions of events that would contaminate our signal, possible acceptance effects in the di-muon mass distribution, related to the choice of the mass selection at generator level, have been studied. Figure 8.2 reports, for the three semi-leptonic samples under study, the scatter plot of the di-muon versus hadron-muon system mass. Due to the small difference between the muon and the pion invariant mass (~ 30 MeV) no significant differences between the two mass distributions have been found, expecting only substantial variations in case of kaons

and protons where the difference in terms of invariant mass is ~ 390 MeV and ~ 833 MeV respectively.



(a) $B^0 \to \pi\,\mu\,\nu$

(b) $B_s^0 \to K\mu\,\nu$



(c) $\Lambda_b^0 \to p\,\mu\,\nu$

Fig. 8.3. *Invariant mass distribution of the di-muon system at generator level for $B^0 \to \pi\,\mu\,\nu$ (left), $B_s^0 \to K\,\mu\,\nu$ (middle) and $\Lambda_b^0 \to p\,\mu\,\nu$ (right) semi-leptonic decays. The distributions obtained applying a hadron-muon mass cut at generator level of 3.0 GeV (blue solid histogram), 3.5 GeV (red solid histogram) and 4.0 GeV (green solid histogram) are superimposed to the one obtained without any mass cut (black solid histogram).*

Figure 8.3 shows the di-muon mass distributions obtained considering three different hadron-muon mass selections at generator level set at 3.0, 3.5 and 4.0 GeV compared to the one obtained without any mass selection. No significant effects on the mass shape are found. Since the lower range of the left sideband is set to 4766 MeV, it was decided to apply a cut at generator level at 3.5 GeV in order to be able, if needed, to extend the left sideband to lower mass values and perform additional checks.

After all these tests, the $B^0 \to \pi\,\mu\,\nu$, $B_s^0 \to K\,\mu\,\nu$ and $\Lambda_b^0 \to p\,\mu\,\nu$ semi-leptonic samples have been finally produced with a statistic of 500 k, 100 k and 100 k generated events respectively.

8.2.2 The "Four-Corners" Background Sample

In order to have a statistically meaningful sample of background MC for the training of the BDT and the background studies for the invariant mass fit, for the 2012 analysis a very large number of events (so far the largest MC production performed by ATLAS for a single analysis) was generated.

Since the necessary statistic for the 2012 analysis is around 7 G events, an inclusive sample (like that used for the 2011 analysis) would have been impossible to produce in a reasonable time scale. Therefore, the specific topologies that dominate the combinatorial background have been selected and produced semi-exclusively. The MC sample was designed to provide a realistic composition of oppositely charged di-muon pairs from different sources (tuning the contribution from individual b- and $\bar{b}$-hadrons, including those from J/ψ decays and from τ leptons, and including non combinatorial sources of muon pairs). On the other hand, because of the uncertainties in the cross sections and the branching ratios, as well as in the generation procedures, no attempt is made to use the MC to draw quantitative conclusions on the background normalisation and shapes. For this purpose, events collected in the sidebands are used.

Allowing only semi-leptonic decays of b hadrons (admixture of $B^{\pm}$, B^0, B_s^0, b-baryons) is not a comprehensive strategy, so the procedure is to consider both primary semi-leptonic decays from b quarks and secondary decays from c quarks able to produce a muon, and combine independent samples of MC with forced semi-leptonic decays of both types, so that the merged samples provide a realistic distribution of oppositely-charged muon pairs in the final state.

Figure 8.4 (left), provides a schematic diagram of the decays of b-hadrons, identifying the sources of muons from primary decays ($b \to c\,\mu^-\,\bar{\nu}$, top diagram) and secondary decays ($c \to s\,\mu^+\,\nu$, bottom diagram). Muon pairs from the same tree (cascade events $b \to c\,\mu^-\,\bar{\nu} \to s\,\mu^+\,\nu\,\mu^-\,\bar{\nu}$ and single-vertex di-muons such as $B \to J/\psi \to \mu^+\mu^- X$) are also shown. The length of each coloured segment is approximately proportional to the corresponding branching fractions.

Muon pairs from decays in the two decay trees are termed *opposite-side* muons and they are shown diagrammatically in figure 8.4 (right). The square is obtained combining muon sources from the b and $\bar{b}$ decay trees: events containing muons from both decay trees populate the regions near the corners of the diagram. The bands connecting the corners are populated by events with one muon, or di-muons originating from the same tree. The total area covered by the four corners is $\sim 4\,\%$ of the full $b\bar{b}$ decay spectrum.

A large sample of $\mu^+\mu^-$ pairs from combinatorial background has been obtained generating events only for the four corners of figure 8.4,

- AA [from A(b)A($\bar{b}$)]: $b \to c \to \mu^+$, $\bar{b} \to \bar{c} \to \mu^-$
- AB [from A(b)B($\bar{b}$)]: $b \to \bar{c} \to \mu^-$, $\bar{b} \to \mu^+$
- BA [from B(b)A($\bar{b}$)]: $b \to \mu^-$, $\bar{b} \to c \to \mu^+$
- BB [from B(b)B($\bar{b}$)]: $b \to \mu^-$, $\bar{b} \to \mu^+$

from which the name "four-corners" is given to this dataset.

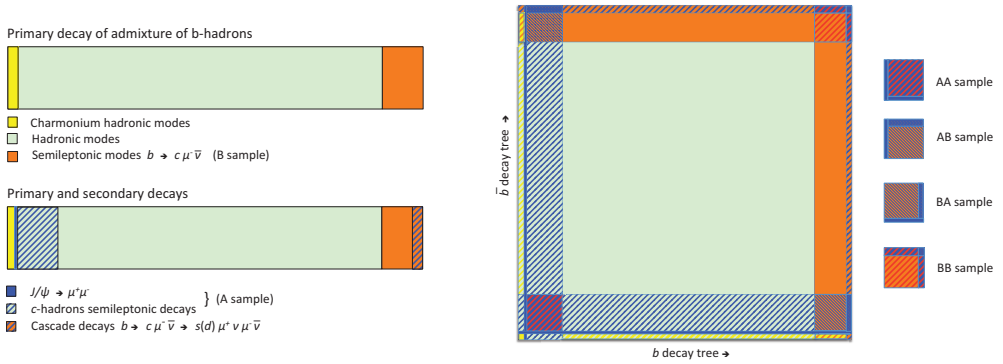


Fig. 8.4. *Left: schematic diagram of the b decay channels: events are classified according to the presence of muons in the finals state, distinguishing primary (b-hadron) and secondary (c-hadron) decays. Right: diagram of the combinations of b and $\bar{b}$ decays: events containing muons from both decay trees are in the four corners, labelled as shown in the legend.*

The scheme described above is implemented using the opportunities for modifying the inclusive decay trees provided by EvtGen. Four samples are generated, and for each of them a dedicated EvtGen configuration file describing the decay trees has been prepared forcing only desired topologies for that sample. For example, in the "BA" sample all hadrons containing a b quark (B^-, $\bar{B}^0$, $\bar{B}^0_s$, and Λ_b) are forced to decay only into final states containing μ^-, i.e. the rest of the decays are switched off (sample B). Similarly, the $\bar{b}$-hadrons (B^+, B^0, B^0_s, and $\bar{\Lambda}_b$) are forced to decay into a c-hadron, which is then forced to a semi-leptonic muon decay (sample A). Additional c-hadrons are decayed according to the standard decay tables.

The correct proportion of muons from decays of B^0, B^0_s, B_u and Λ_b is naturally obtained by the standard fragmentation scheme implement in Pythia, and by the similar values of their semi-leptonic branching fractions. For c-hadrons, the larger branching fraction for D^+ is correctly reproduced by allowing a fraction of D^0, D_s, Λ_c to decay non-leptonically, so that the sums of the allowed channels correspond to the D^+ semi-leptonic branching fraction.

Effects of B^0 and B^0_s oscillations, as well as decays $b \to c\bar{c}s(d)$ with $\bar{c} \to \mu^- X$ are properly handled in the simulation, and contribute significantly to the number of opposite-charge pairs in the samples "AB" and "BA".

Muon pairs from the same decay vertex (*e.g.* $B \to J/\psi\, X \to \mu^+\mu^- X$), cascade associations ($b \to c\,\mu^-\bar{\nu} \to s(d)\,\mu^+\mu^-\nu\,\bar{\nu}$), and muons from the decay of τ leptons are present in the four-corners MC sample, although their normalisation, compared to an unbiased simulation, is not as much enhanced as that of opposite side semi-leptonic decays. They represent a small additional contribution to the dominant combinatorial background from opposite-side semi-leptonic decays. Similarly, B_c mesons are produced in the fragmentation of b ($\bar{b}$) quarks, but in this case the semi-leptonic decays are not enhanced.

The relative weights of the four samples are calculated taking into account the branching fractions and the efficiencies of the generator-level selections. The production factors were determined comparing the yield of events in the different classes to those obtained in an full inclusive MC.

8.2.2.1 Data Driven Weights for the "Four-Corners" MC

The behaviour of the combinatorial part of the four-corners MC and the potential inaccuracies present in the simulation can be checked using the 2012 sideband data.

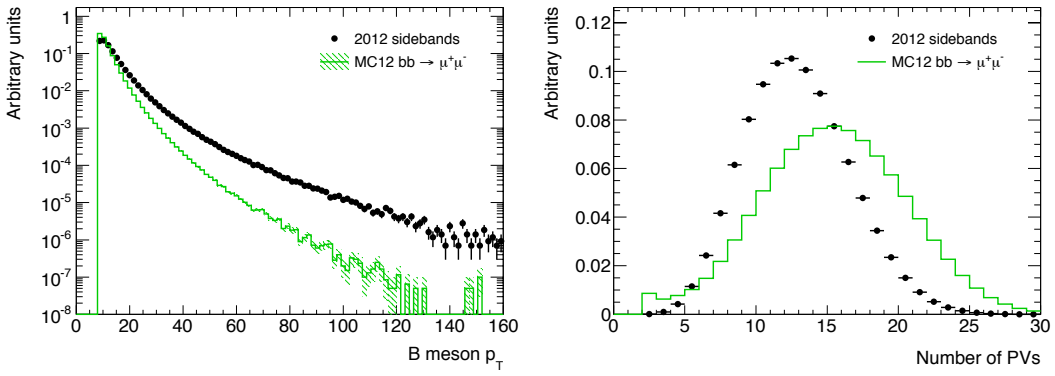


Fig. 8.5. *Distributions of the B meson p_T (left) and number of primary vertices (right) for the four-corners MC (green) and sidebands data (black).*

The data-MC disagreement present in some of the discriminating variables used for the continuum background rejection would cause the training of the BDT to be substantially suboptimal. Part of these discrepancies are related to an inaccurate pile-up[2] and B kinematic simulation in the four-corners MC. Therefore, a set of data-driven weights, evaluated using the same iterative method explained in section 5.5.4, has been introduced. The weights have been derived using the $B^0_s \to \mu^+\mu^-$ sideband data. Figure 8.5 reports the distribution of the p_T of the B meson and the number of primary vertices before the re-weighting procedure, showing the data-MC disagreement.

Figure 8.6 reports the data-driven weights, as a function of the p_T of the B meson and the number of primary vertices, obtained with the above procedure. Since the weights have been evaluated using the full MC and sidebands statistic, in order to reduce the correlation between data and four-corners MC, the p_T weights have been parameterised using a 3^{rd} order polynomial function $w\,(p_T) = a_0 + a_1 \cdot p_T + a_2 \cdot p_T^2 + a_3 \cdot p_T^3$ where $a_0 = -0.55 \pm 0.01$, $a_1 = 0.133 \pm 0.003$, $a_2 = -0.00007 \pm 0.00016$ and $a_3 = 0.000016 \pm 0.000003$.

The result of the re-weighting procedure is reported in figure 8.7 that shows, as sanity check, the data-MC comparison for both the p_T of the B meson candidate and the number of primary vertices.

[2]This is due to the MC release used to produce the background samples.

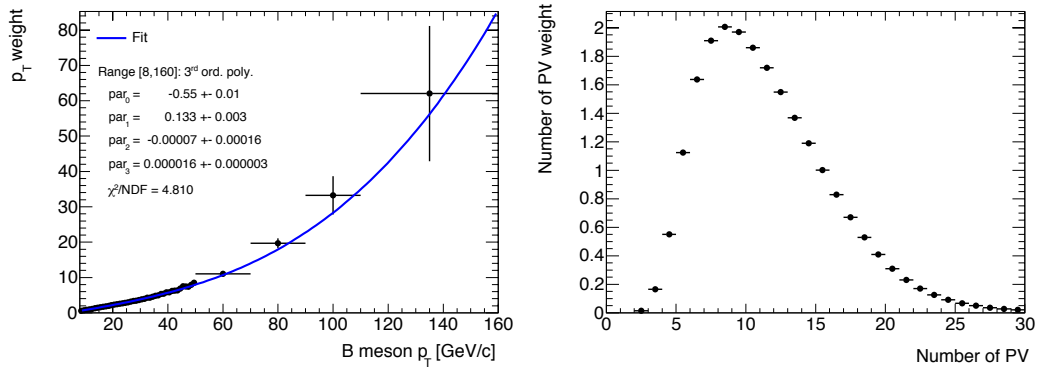


Fig. 8.6. *Data driven weights as a function of the B meson p_T (left) and the number of primary vertices (right). The p_T weights have been parameterised using a 3^{rd} order polynomial function.*

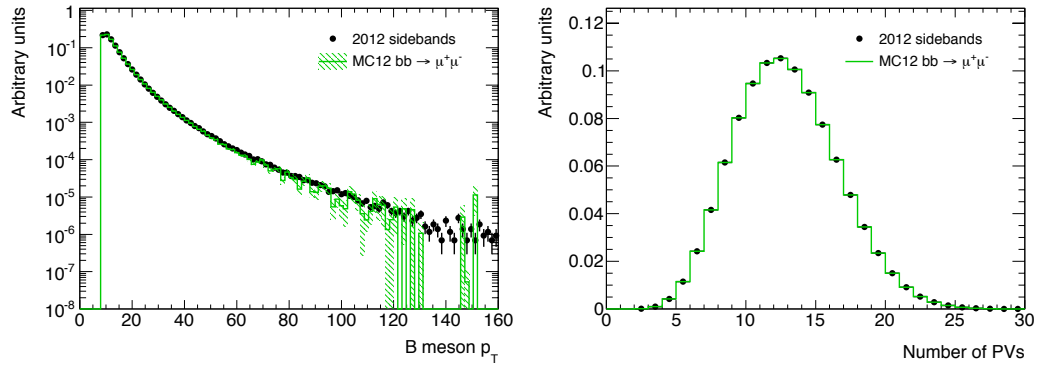


Fig. 8.7. *Distributions of the B meson p_T (left) and number of primary vertices (right) for the four-corners MC (green) and sidebands data (black) after the re-weighting procedure.*

8.3 Event Selection

The blinding region for the $B_{s,d} \to \mu^+\mu^-$ signal channel has been reduced (with respect to the 2011 analysis) to a width of 360 MeV, $[-200, +160]$ MeV around the B_s^0 mass corresponding to the mass region [5166-5526] MeV. This allows to have a wider left sideband that can be used to precisely constrain the shape of the low-mass background on data. As it will be discussed in section 8.4, the 2012 analysis benefits also of the new mass definition that considers the combined (ID-MS) muons in its computation.

In addition, the determination of the event primary vertex (PV) associated to the B candidate has been revisited with respect to the 2011 analysis. The previous analyses performed the association choosing the PV closest to the decay vertex along the z direction. However, in presence of higher pile-up levels, this method can select the

wrong PV, in particular for B mesons produced at high rapidity, where the displacement between the true primary vertex and the decay one is oriented close to the detectors axis. A new method has been developed, in which the backward extrapolation of the B momentum from the decay vertex is considered, and the associated PV is chosen as the one with the shorter separation, along z, from the point of the closest approach (POCA) of the B momentum extrapolation to the beam line (see figure 8.8).

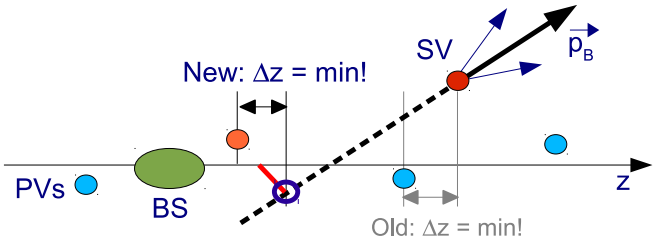


Fig. 8.8. *Sketch of the primary vertex association method. This is not to scale and for illustration purposes only.*

The baseline selection follows essentially the one reported in section 5.4. The condition on the common vertex, χ^2/NDF, has been released from 2 to 6 since in the 2012 analysis this variable is included in the BDT selection. In addition, the cuts on the number of Pixel, SCT and TRT hits for only the muon tracks have been modified following the guidelines of the Muon Performance (MP) working group, now requiring

- > 0 (Pixel hits + crossed dead Pixel sensors)
- > 4 (SCT hits + crossed dead SCT sensors)
- if $0.1 < |\eta| < 1.9$:
  (TRT hits + TRT outliers) > 5 and
  (TRT outliers) $< 0.9 \cdot$ (TRT hits + TRT outliers)
- < 3 (Pixel + SCT) holes

The additional selection for the background reduction, described in section 6.1.3, has been tested using 2012 MC and sideband data and showed the same performances, in terms of signal efficiency and background reduction, obtained on 2011 datasets. Therefore, the same selection ($\alpha_{2D} < 1.0$, $\Delta R < 1.5$ and $L_{xy} > 0$) has been applied to all 2012 samples after the baseline selection described above.

8.4 New 2012 Mass Definition and Parameterisation

The mass resolution is strictly related to the momentum resolution of the muons. Since the muons originating from the B_s^0 can have different η, they cross different amount and type of detectors. For that reason, it is useful to sub-divide the detector in three different regions (see figure 8.9). "barrel" (B) in which a track crosses only

the ID barrel layers ($0 < |\eta| < 1.0$), "transition" (T) ($1.0 < |\eta| < 1.5$) and "end-cap" (E) ($1.5 < |\eta| < 2.5$).

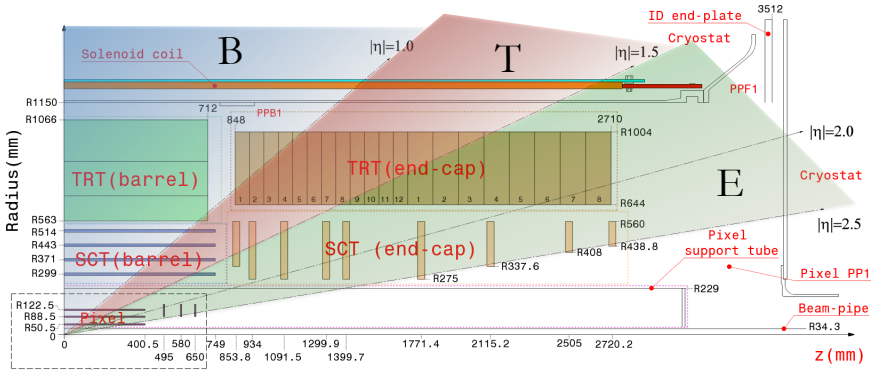


Fig. 8.9. *Mass resolution segmentation as a function of η.*

In the past analyses the di-muon invariant mass has been evaluated considering only track information from the ID, while for 2012 data the calculation has been performed using the combined-muon tracks (see section 2.2.5). Including the muon spectrometer information improves the mass resolution by about 20 MeV when muons are reconstructed in the end-caps.

Starting from the above definition, we can separate all our events in three different regions identified as a function of the maximum $|\eta|$ of the two muons ($\eta_{\max}$)[3]. When $\eta_{\max}$ is in the ranges [0, 1.0]/[1.0, 1.5]/[1.5, 2.5], the average signal mass resolution is approximately 61/78/107 MeV respectively. The three regions correspond to 51/24/25 % of the total number of events respectively. Since the effects of the different mass and momentum resolution of the ATLAS detector are separated, in each region the invariant mass distribution can be fitted using only a single Gaussian and the results (for the ID-based and combined-muons mass definition) are reported in figure 8.10.

In the analysis performed on 2011 data the possibility of separately optimising the selection for the three categories of events has been explored, in order to take advantage of the improved background rejection achievable for the classes with better resolution. Since no enhancement in the signal sensitivity was found, the upper limit was extracted using the whole $\eta_{\max}$ range.

For the 2012 analysis this procedure has been studied in more detail. One can split data into the three different regions, perform the invariant mass fit in each of them simultaneously (modelling the signal with a single Gaussian) and then combine the results obtained to get the total number of signal events. An alternative strategy would be to parameterise the mass resolution as a function of $\eta_{\max}$ and then fit the signal MC using a single Gaussian with a width depending on the resolution region of the muons. The $\eta_{\max}$ range on signal MC has been subdivided in 12 bins and for each of them a mass fit using a single Gaussian has been performed to extract the corresponding resolution. The results, for ID-only and combined-muons mass definition, are reported

[3]Similarly, the $|\eta|$ of the B meson can be used.

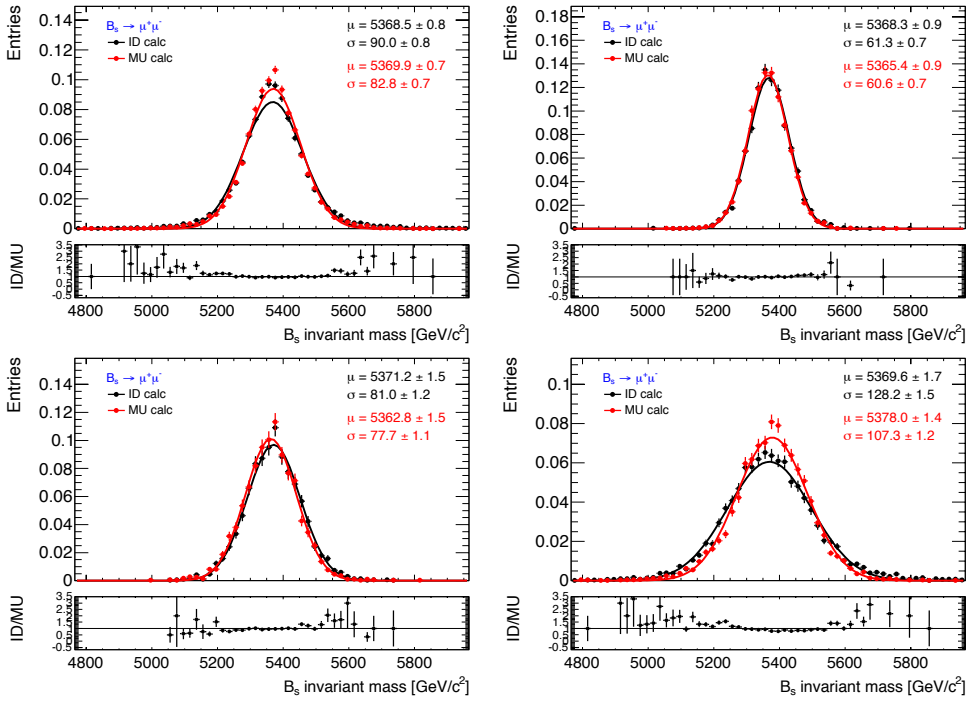


Fig. 8.10. *Comparison of the di-muon invariant mass obtained with ID-based (black dots) and combined-muons (red dots) mass calculation (a single Gaussian fit is superimposed on the distributions). Top-left: all the events in the full η_{max} range. Top-right: events in the first resolution zone. Bottom: events in the second (left) and third (right) resolution regions.*

in figure 8.11 and have been parameterised using respectively a 4^{th} order and 3^{rd} order polynomial function.

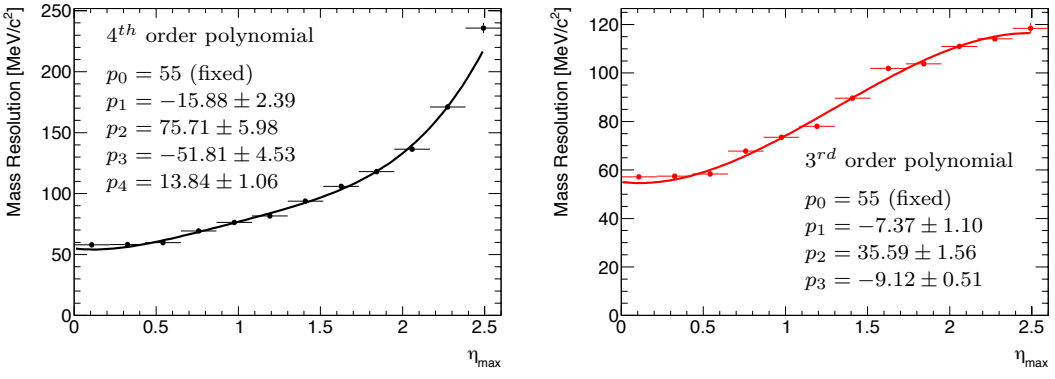


Fig. 8.11. *Mass resolution parameterisation as a function of η_{max} for ID-only (left) and combined-muons (right) mass calculation.*

A complementary strategy that has been investigated is the one that uses a single

Gaussian model with a resolution depending on the per-event vertex mass error δm

$$\mathcal{S}(\boldsymbol{m};\mu,s,\delta m) = e^{-\frac{(\boldsymbol{m}-\mu)^2}{2\,(s\cdot\delta m)^2}} \quad , \tag{8.2}$$

where s is a resolution scaling parameter independent of $\eta_{\max}$. A similar technique has also been adopted in the 2011 analysis for the mass fit of the reference channel (see section 5.7), where the per-event mass error was modelled using a Gaussian kernel p.d.f.. Figure 8.12 reports the mass error distributions for ID-only and combined-muons mass definition in the full $\eta_{\max}$ range as well as in the three resolution regions. Due to a very complex pattern, the error distribution can not be easily modelled in the full range. However, splitting the events in the three categories the error assumes a Gaussian-core with a power-law right-end tail shape. As was done for the mass resolution, the $\eta_{\max}$ range has been split in 12 bins, and for each of them the corresponding error distribution has been fitted using a Crystal Ball function. The results of the fits in each bin, for the two mass calculations, are reported in figure 8.13 that shows also the behaviour of the mean of the Crystal Ball as a function of $\eta_{\max}$. The improved resolution in the end-cap region is highlighted by the increased separation between the two distributions in the last bins.

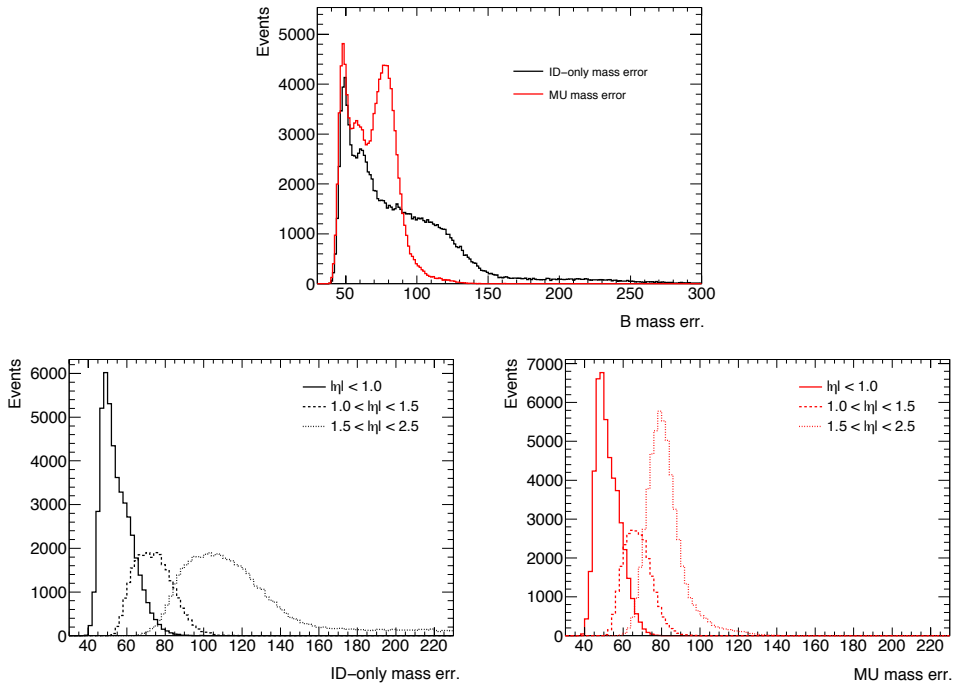
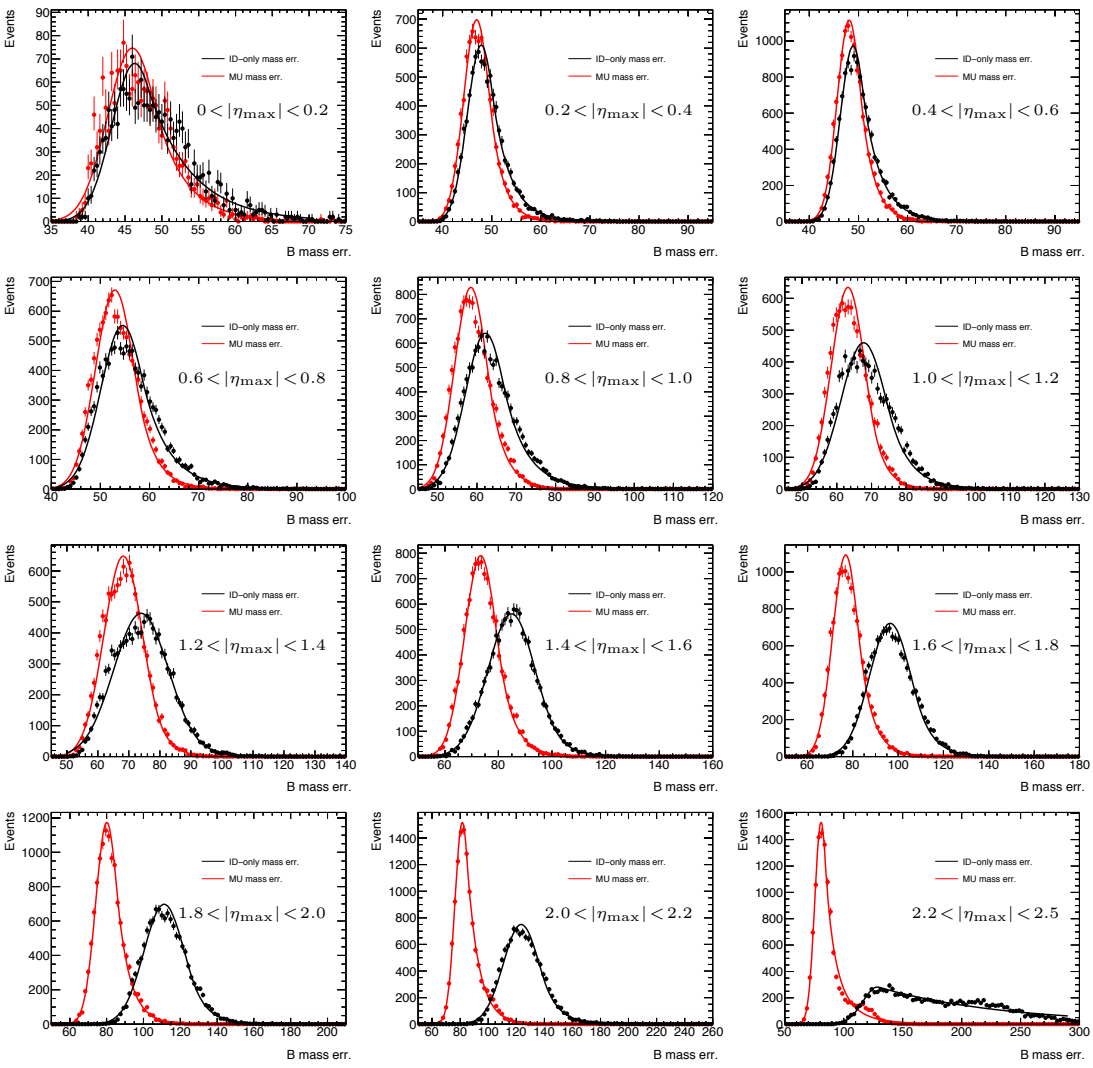
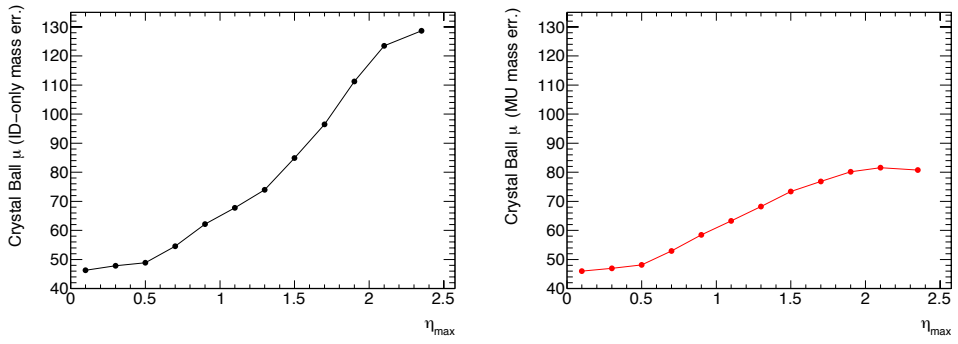


Fig. 8.12. *Top: di-muon vertex mass error in the full η_{max} range for ID-only (solid black line) and combined-muons (solid red line) mass definition. Bottom: di-muon vertex mass error in the three resolution regions for ID-only (left) and combined-muons (right) mass definition.*

The mass resolution and the per-event mass error parameterisations have been tested using the final version of the invariant mass fit discussed in section 8.10. Toy



(a)



(b)

Fig. 8.13. ***(a)****: di-muon vertex mass error in 12 η_{max} bins with the Crystal Ball fit superimposed on the distributions. Black dots (and solid black lines) refer to the ID-only mass definition while red dots (and solid red lines) to the combined-muons one.* ***(b)****: behaviour of the mean of the Crystal Ball as a function of η_{max} for ID-based (left) and combined-muons (right) mass definition.*

MC experiments proved how with both strategies there is no real gain in terms of signal significance with respect to the simple fit performed using a double Gaussian and therefore, for sake of simplicity, the latter has been used to extract the signal events from 2012 data. Nonetheless, this strategy could be further investigated and developed using Run 2 data.

8.5 Trigger Studies

In 2012, the trigger `2mu4T`, seeded at L1 by `2MU4` and already used in the 2011 analysis (see section 5.4), was prescaled at the beginning of each data taking run in order to avoid exceeding the sustainable rate. This effectively reduces the amount of data that could be selected by this trigger stream from 20.3 fb^{-1} to 16 fb^{-1}. The left plot in figure 8.14 shows the behaviour, during one data taking run, of the trigger `EF_2mu4T_Bmumu_Barrel` [4] (that is not prescaled) compared to `EF_2mu4T` that is prescaled at the beginning of each run. To avoid losing 20 % of signal events, other triggers

- `EF_2mu4T_Bmumu`
- `EF_2mu4T_Bmumu_Barrel`
- `EF_mu4Tmu6_Bmumu`
- `EF_2mu4T_Bmumu_BarrelOnly`
- `EF_mu4Tmu6_Bmumu_Barrel`
- `EF_2mu6_Bmumu`
- `EF_3mu4T`

can be included in the selection in order to (partially) recover the luminosity lost with the prescaling (for more details about the B-physics triggers see section 2.3.2).

Considering the above triggers, it is possible to recover those events not saved by the default `2mu4T`, effectively counting the number of events that do not pass `2mu4T`, but pass another specific trigger with the following logic:

- `EF_2mu4T_Bmumu`
- `EF_2mu4T_Bmumu_Barrel && !(EF_2mu4T_Bmumu)`
- `EF_mu4Tmu6_Bmumu && !(EF_2mu4T_Bmumu_Barrel || EF_2mu4T_Bmumu)`
- and similarly for the other triggers.

Since roughly 98 % of the events on data passed the first three triggers, considering only these is an acceptable simplification. This is also clear in the right plot of figure 8.14 that shows the number of events selected by the various triggers in 2012 data. The logic OR of all triggers shows how the first three are able to provide essentially the whole statistic.

Restricting the analysis to the three main triggers, the relative fraction of events normalised to the total number of events in the sideband data and in the exclusive

[4] The trigger `2mu4T_Bmumu_Barrel` requires to have at least one muon to be triggered in the barrel of ATLAS, while the trigger `EF_2mu4T_Bmumu_BarrelOnly` require to have both muons in the barrel.

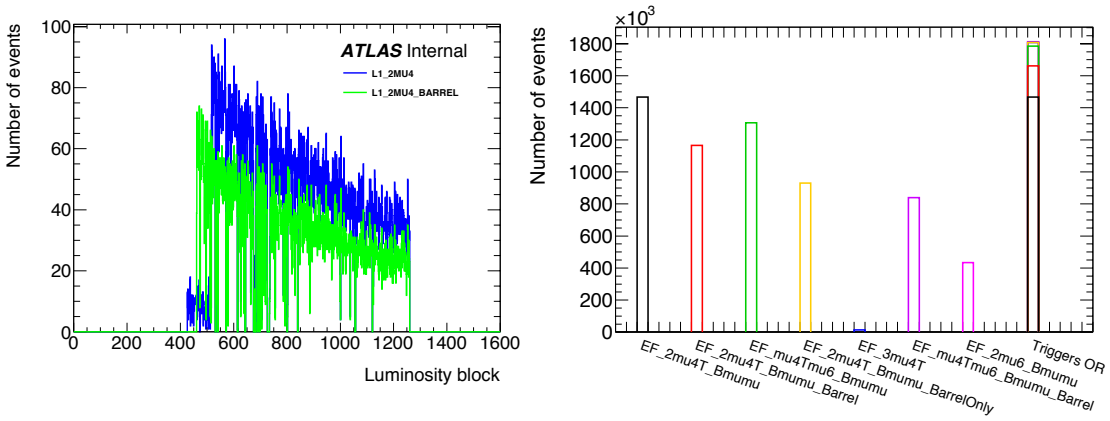


Fig. 8.14. *Left: number of events triggered at L1 as 2MU4 (solid blue line) and 2MU4_Barrel (solid green line) as function of the luminosity block (∼ 2 minutes of data taking) in a run in period I (end of October - beginning of November 2012), where the trigger L1_2MU4 was downscaled at the beginning of the fill. Right: number of events selected by each relevant EF di-muon trigger in the whole 2012 data.*

Trigger category	**Relative event yield**			
	DATA		**MC**	
	di-muon	$B^{\pm}$	**di-muon**	$B^{\pm}$
N_1	6.6 %	7.5 %	10.0 %	10.4 %
N_2	22.8 %	15.3 %	18.0 %	24.5 %
N_3	70.6 %	77.2 %	72.0 %	65.1 %

Table 8.2. *Signal $B_s^0 \to \mu^+\mu^-$ and reference channel $B^{\pm} \to J/\psi K^{\pm}$ relative fraction of events for the three trigger categories defined in the text, evaluated on sideband data and MC events.*

MC samples have been checked (both for the signal and the reference channel). The schema to be applied to data for defining the relative fractions is characterised to better isolate the specific topological differences:

- N_1: `EF_2mu4T_Bmumu && !(EF_2mu4T_Bmumu_Barrel || EF_mu4Tmu6_Bmumu)`
- N_2: `EF_2mu4T_Bmumu_Barrel && !(EF_mu4Tmu6_Bmumu)`
- N_3: `EF_mu4Tmu6_Bmumu`

where N_1, N_2 and N_3 are the three trigger categories considered in the master formula (equation 8.1). Table 8.2 shows the relative fraction of events for the three trigger categories normalised to the total number of events evaluated on data and MC for the signal and the reference channel. From the ratios of the relative fraction, we can extract the weights to correct the MC to the data relative abundances.

Figure 8.15 reports the corresponding p_T and η distributions of the events highlighting the differences in the topology of the three categories.

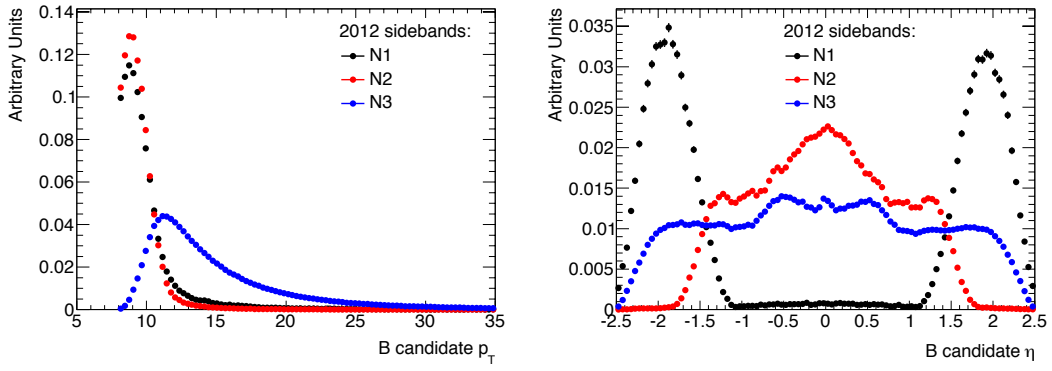


Fig. 8.15. *Distributions of the p_T (left) and η (right) of the B candidates obtained for the three different trigger categories.*

8.6 Studies on the Muon Fake Rates

One of the most problematic backgrounds in the $B_s^0 \to \mu^+\mu^-$ analysis is represented by the charmless two-body B decays referred to as $B \to hh'$, h being a charged K or π. This background is topologically identical to, and peaks under, the signal (see figure 8.16). The only handle we can exploit is the muon identification capability of the ATLAS detector. For these decays to feed into our events, the charged K or π has to be misidentified as a muon. In the previous analysis [106], the muon fake rate was around 3.9 ‰ for kaons and 2.1 ‰ for pions. For the 2012 data, dedicated studies, based on a multivariate analysis, were performed in order to reduce the fake rate, thus decreasing the signal contamination from these background events.

The study of the muon fake rates has been performed on two MC samples of signal ($B_s^0 \to \mu^+\mu^-$) and charmless two-body decays ($B \to hh'$). These samples have been produced with full GEANT4 simulation in order to accurately describe the hadron responses in the calorimeters. The preselection described in section 8.3 is applied to the events entering this study.

Table 8.3 shows the misidentification fraction for protons, kaons, and pions after the preliminary cuts, obtained through the full simulation of the decays of b-hadrons to pairs of charged, long-lived hadrons, in which one of the two is erroneously identified as a combined muon. The vast majority of muons originate from decays in flight of kaons and pions. A dedicated MC production, in which the propagation and the interactions of the hadrons in the GEANT4 simulation are reported in detail, has shown that decays in flight occur in 97 % (92 %) of the cases when pions (kaons) are mistaken as muons. The remaining cases correspond to punch-through, and the small fraction of such events explains the negligible contribution of protons and antiprotons to the

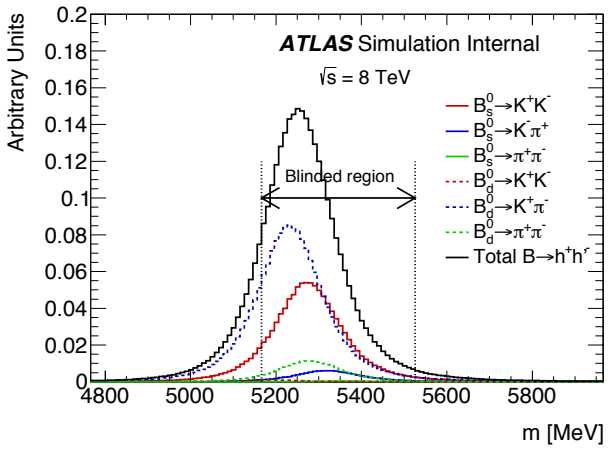


Fig. 8.16. *Invariant mass distribution of the peaking background components $B \to hh'$.*

total hadron misidentification. The requirement of matching a muon candidate to a muon trigger object reduces further the number of hadron misidentified as muons by a factor equal to 0.582 ± 0.015, the same for kaons and pions. The association is based on the angular separation $\Delta R = \sqrt{\Delta\phi^2 + \Delta\eta^2}$.

Misidentified hadron	MC sample	Fake rate (after preliminary cuts)
$p(\bar{p})$	$\Lambda_b \to ph$	$(3.30 \pm 0.71)\ \times 10^{-5}$
K/π	$B \to hh'$	$(3.01 \pm 0.03)\ \times 10^{-3}$
	$\Lambda_b \to ph$	$(3.31 \pm 0.07)\ \times 10^{-3}$

Table 8.3. *Misidentified hadron, MC samples (with full simulation) used for the studies of fake muons, and fraction of hadrons identified as muons after the preliminary selection.*

After the preliminary selection and the trigger match, the rejection of fake muons includes a BDT selection (fake-BDT). In collaboration with the MP working group, a number of discriminating variables tailored to our specific decay topology has been selected in order to build a multivariate variable capable of discriminating between true and fake muons. Some of the variables used in the BDT are the energy of the muon in the MS, the energy loss in the calorimeter, the muon track fit χ^2, the cluster energy E_T in a ΔR cone around the muon trajectory and the number of track hits in the MDT, RPC and TGC (see section 2.2.5).

Table 8.4 shows the reduction in the fraction of fake muons obtained with a fake-BDT threshold corresponding to a muon selection efficiency equal to 95 %. The fake fraction would be further reduced by a factor ~ 0.8 if the selection is tuned to 90 % muon efficiency. Table 8.5 shows the overall performance of the fake-muons reduction

procedure, showing the fraction of hadrons misidentified as muons after preliminary cuts, adding the trigger match and the fake-BDT selection corresponding to a muon selection efficiency equal to 95 %.

Hadron	Fake-muons reduction factor
$K^{\pm}$	0.376 ± 0.007
$\pi^{\pm}$	0.366 ± 0.010

Table 8.4. *Fake-muons reduction factors with statistical uncertainty obtained with the fake-BDT selection corresponding to 95 % muon efficiency.*

Hadron	+ preliminary selection	+ trigger	+ fake-BDT selection
K^-	3.60×10^{-3}	2.07×10^{-3}	$(0.76 \pm 0.03) \times 10^{-3}$
K^+	4.40×10^{-3}	2.63×10^{-3}	$(1.01 \pm 0.03) \times 10^{-3}$
π^-	2.02×10^{-3}	1.16×10^{-3}	$(0.44 \pm 0.02) \times 10^{-3}$
π^+	2.06×10^{-3}	1.21×10^{-3}	$(0.42 \pm 0.02) \times 10^{-3}$
Average	3.09×10^{-3}	1.81×10^{-3}	$(0.67 \pm 0.01) \times 10^{-3}$

Table 8.5. *Cut flow of hadron misidentification fraction. The fake-BDT selection is tuned for 95 % muon efficiency and the error is the statistical uncertainty after all cuts.*

The final fake rates obtained with this optimised selection are lower than 1 ‰ for both kaons and pions. The proton fake rate, after the baseline and trigger selection, is already at the order of 10^{-5}.

8.7 Rejection of the Combinatorial Background

The $B_s^0 \to \mu^+\mu^-$ signal events are separated from the continuum background using a discriminating variable obtained by a MVA tool (TMVA) embedded in the ROOT framework as was done for the 2011 analysis. Detailed studies have been performed analysing new discriminating variables and BDT configurations in addition to those used in 2011. The previous selection has been improved and now benefits from a higher discriminating power provided by 22 variables, described in table 8.6, and of an optimised tuning of the BDT algorithm.

The BDT training is done on MC using the new four-corners background and the signal $B_s^0 \to \mu^+\mu^-$ events. The weights described in section 8.2.2.1 have been applied to the four-corners, while the signal $B_s^0 \to \mu^+\mu^-$ have been re-weighted using the GL corrections and DD weights as for the 2011 analysis. In order to exclude the B-like components present in the four-corners, only opposite-side muons events (see section

Variable	Description	Comments
Isolation $I_{0.7}^{\mu^\pm}$	Ratio of $\lvert\overrightarrow{p}_T^{\mu^\pm}\rvert$ and the transverse momenta of all tracks with $p_T > 0.5$ GeV within a cone $\Delta R < 0.7$ from the $\mu^\pm$ direction, excluding B decay products (2 input variables, one per muon)	new
$N_{trk}^{Iso_{07}^\pm}$	Number of tracks in the $\mu^\pm$ isolation cone (2 input variables, one per muon)	new
L_{xy}	Scalar product in the transverse plane of $(\Delta\overrightarrow{x} \cdot \overrightarrow{p}^B)/\lvert\overrightarrow{p}_T^B\rvert$ where $\Delta\overrightarrow{x}$ is the vector between PV and B decay vertex	used in 2011
Isolation $I_{0.7}$	Ratio of $\lvert\overrightarrow{p}_T^B\rvert$ and the transverse momenta of all tracks with $p_T > 0.5$ GeV within a cone $\Delta R < 0.7$ from the B direction, excluding B decay products	improved definition
$\lvert\alpha_{2D}\rvert$	Absolute value of the angle in the transverse plane between $\Delta\overrightarrow{x}$ and $\overrightarrow{p}^B$	used in 2011
$\lvert\alpha_{3D}\rvert$	Absolute value of the angle between $\Delta\overrightarrow{x}$ and $\overrightarrow{p}^B$	new
P_L^{min}	Minimum momentum of the two muon candidates along the B direction	used in 2011
p_T^B	B transverse momentum	used in 2011
p_T^B sig.	B transverse momentum significance	new
$\log(\chi_{xy}^2)$	Significance of the separation between production (PV) and decay (SV) vertices, $\Delta\overrightarrow{x}^T \cdot \left(\sigma_{\Delta\overrightarrow{x}}^2\right)^{-1} \cdot \Delta\overrightarrow{x}$, in the transverse plane (x, y)	used in 2011
ΔR	Angle $\sqrt{(\Delta\phi)^2 + (\Delta\eta)^2}$ between $\Delta\overrightarrow{x}$ and $\overrightarrow{p}^B$	used in 2011
$\lvert d_0\rvert^{max}$ sig., $\lvert d_0\rvert^{min}$ sig.	Absolute values of the maximum and minimum impact parameter in the transverse plane of the B decay products relative to the primary vertex, divided by the relative uncertainty	new
DOCA_{xtrk}	DOCA of the track closest ("xtrk") to the B vertex (x-y plane)	new
N_{trks}^{close}	Number of ("close") tracks with $\ln(\chi^2) < 1$ where χ^2 is a test of association of a track to the reconstructed B vertex. The tracks associated to pile-up vertices are excluded	new
$\Delta\phi(\mu\mu)$	difference in ϕ between the two muons	new
$\text{DOCA}_{\mu\mu}$	DOCA of the two ID tracks forming the B vertex (x-y plane)	new
χ_B^2/NDF	χ^2/NDF of the B vertex	new
$\chi_{\mu,xPV}^2$	minimum χ^2 between the muons and any other PV	new
IP_B^{3D}	3-dimensional impact parameter (POCA) of the B candidate	new

Table 8.6. *Description of the 22 discriminating variables used for the combinatorial background rejection.*

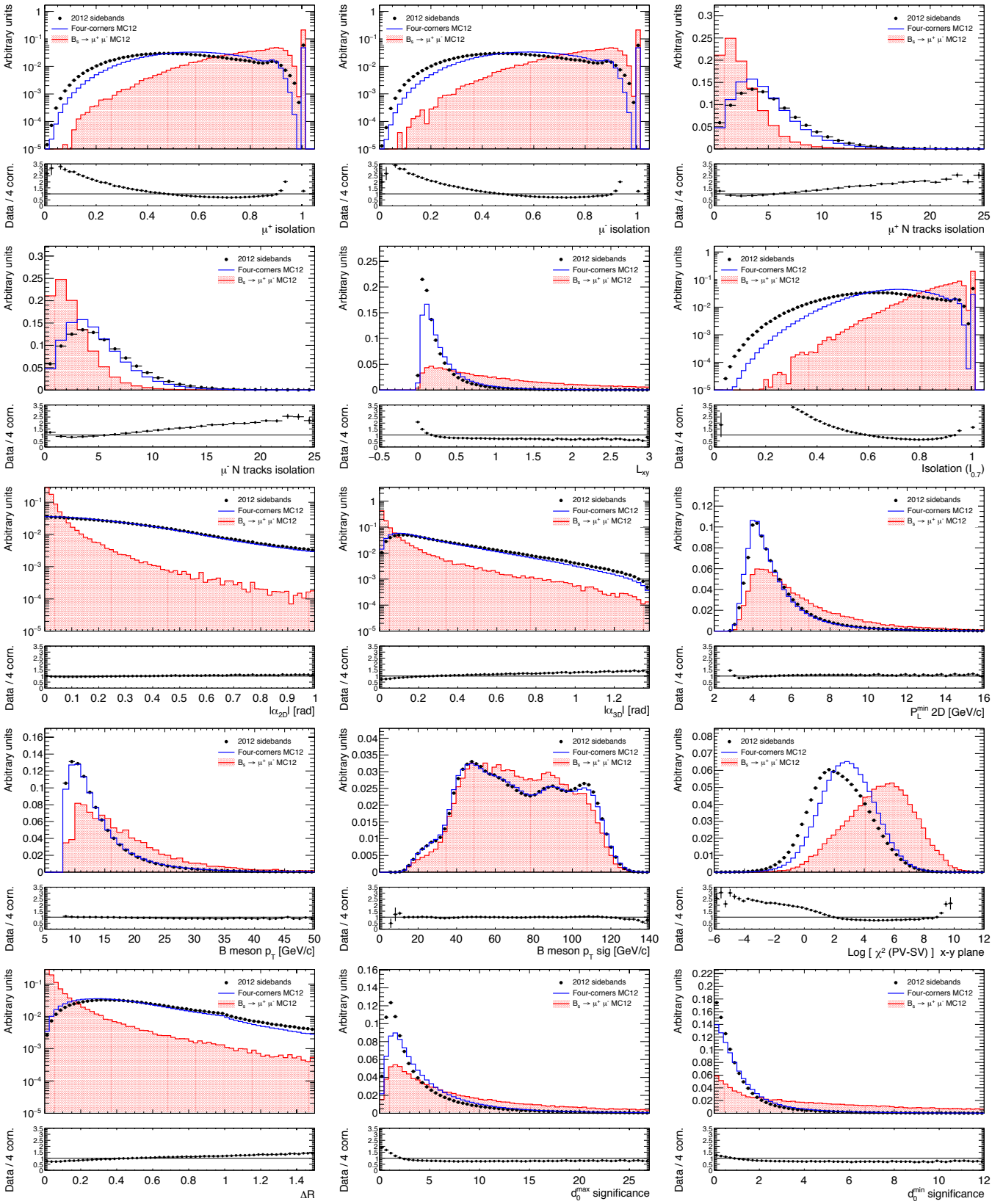


Fig. 8.17. *Distributions of the discriminating variables used in the 2012 analysis. The black dots represent the 2012 sideband data, the blue solid histogram is the four-corners re-weighted MC. For shape comparison only, the signal MC is also shown (red-filled solid histogram). For each plot it is also reported the ratio between sideband data and MC continuum distributions.*

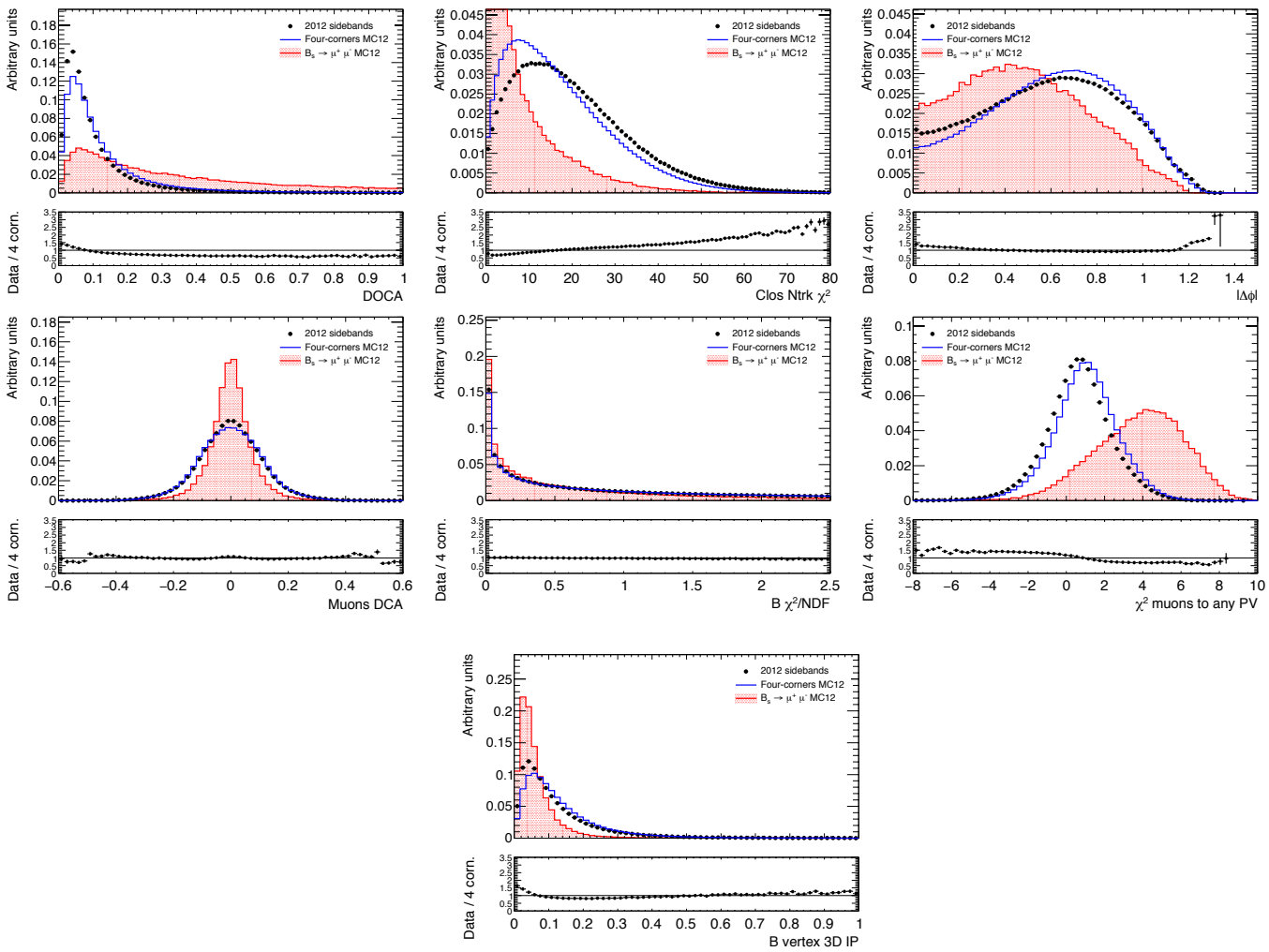


Fig. 8.18. *Distributions of the discriminating variables used in the 2012 analysis. The black dots represent the 2012 sideband data, the blue solid histogram is the four-corners re-weighted MC. For shape comparison only, the signal MC is also shown (red-filled solid histogram). For each plot it is also reported the ratio between sideband data and MC continuum distributions.*

8.2.2 and 8.10.2) have been used for the training. Figures 8.17 and 8.18 show the comparison between 2012 sideband data and four-corners re-weighted MC for all the 22 discriminating variables.

Figure 8.19 reports the comparison of the BDT output distributions for 2012 signal re-weighted MC, four-corners re-weighted MC and sideband data obtained with the BDT configuration used for 2011 analysis and with the new one. The ROC curve for the two different classifiers applied on data 2011 is also shown, proving the significantly higher background rejection of the new BDT configuration obtained for the 2012 analysis.

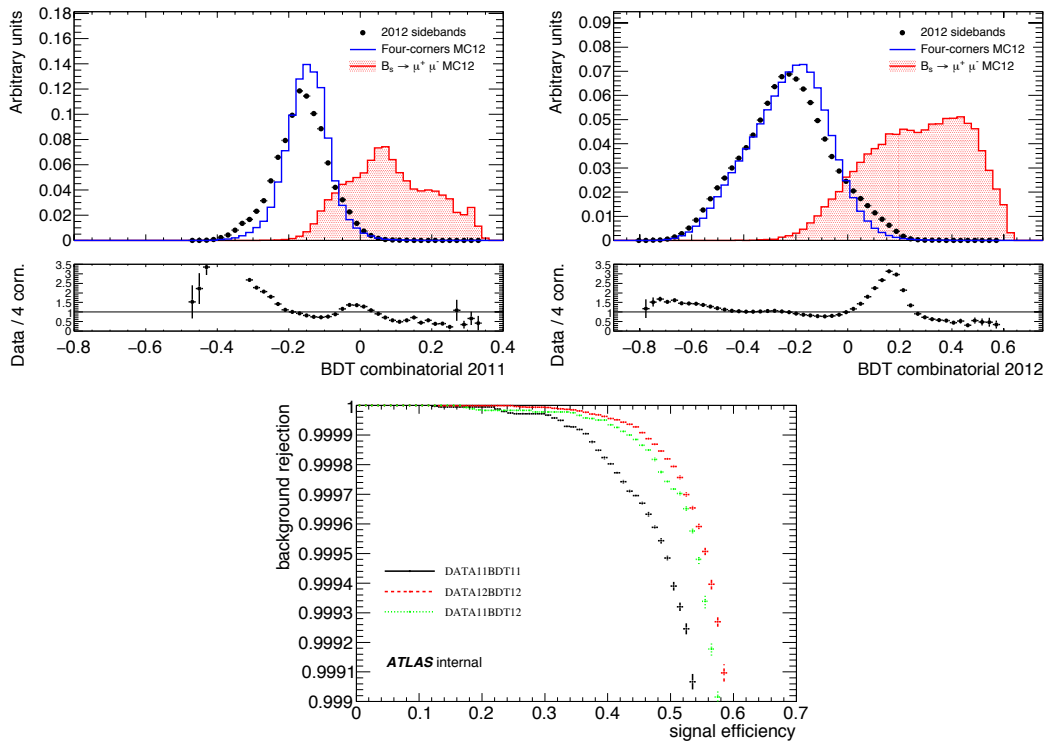


Fig. 8.19. *Top: BDT output distributions for sideband data (black dots), four-corners reweighted MC (blue solid histogram) and signal MC (red-filled solid histogram) obtained with the BDT used in the 2011 analysis (left) and with the new one trained on 22 discriminating variables (right). Bottom: ROC curve obtained with 2012 BDT applied on both 2011 and 2012 sideband data, and with 2011 BDT applied on 2011 sideband data, showing the higher performance of the new classifier.*

8.8 Yield Extraction for the Reference Channel

The $B^{\pm}$ yield for the reference channel is extracted using an un-binned maximum likelihood fit to the mass distribution. A similar fit was used in the 2011 analysis.

As already discussed in section 5.7 for 2011 data, in the $B^{\pm}$ invariant mass distribution the signal is quite evident, but with contributions from at least three background categories (the left-hand plot in figure 8.20 reports the mass distribution obtained with 2012 data). On the left of the signal peak we find the partially reconstructed B decays (e.g. $B^{+/0} \to K^{*+/0}J/\psi$, $B^{+} \to K^{+}\chi_{c1,2}$) where one or more of the final state particles are missed in the reconstruction. On the right side, it is expected a contribution from the reflection of the Cabibbo suppressed $B^{\pm} \to J/\psi\pi^{\pm}$ decay with the assignment of the kaon mass to the final state pion. Finally, the combinatorial background, mostly composed of $b\bar{b} \to J/\psi X$ events, spans the whole mass range and consists of random combinations of a track with a di-muon pair from the J/ψ. For the

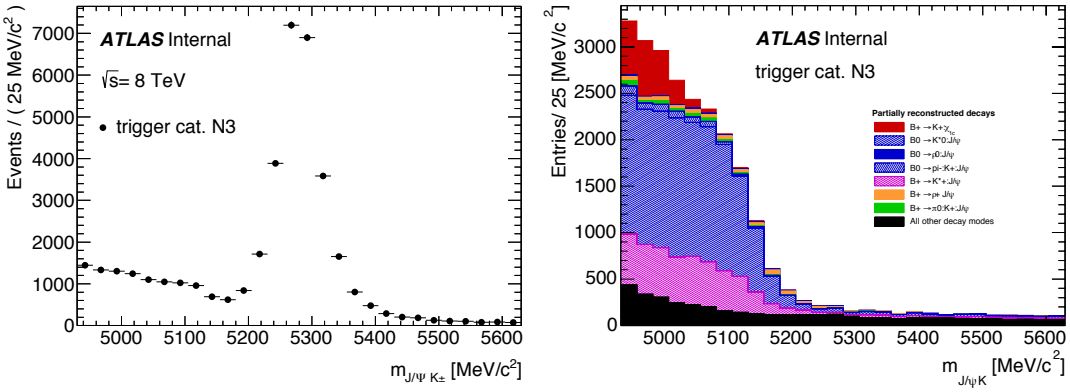


Fig. 8.20. *Left: $J/\psi K^\pm$ invariant mass distribution for all $B^\pm$ candidates for the trigger category N_3 in 2012 data. Right: partially reconstructed B decays contributing to the background as described by MC.*

extraction of the $B^\pm$ yield the following event categories have been defined:

- $N_{J/\psi K^\pm}$: number of $B^\pm \to J/\psi K^\pm$ events (signal events for this fit);
- $N_{J/\psi \pi^\pm}$: number of $B^\pm \to J/\psi \pi^\pm$ events;
- $N_{p.r.}$: number of partially reconstructed events;
- N_{comb}: number of combinatorial events.

Table 8.7 shows the signal candidates, in the mass region $4930 - 5630$ MeV, passing the pre-selection and the additional cuts. The events are separated into the three trigger categories as defined in section 8.5 and these are the candidates entering into the invariant mass fits.

Category	Candidates
N_1	4631
N_2	11677
N_3	110191

Table 8.7. *Signal candidates entering the $B^\pm \to J/\psi K^\pm$ fit for each trigger category.*

The fit procedure is based on an un-binned extended maximum likelihood fit to data, separated in the three categories, with simultaneous inclusion of three MC samples. The MCs are introduced to shape accurately the signal as well as the most critical background components: the partially reconstructed modes and the mis-reconstructed $J/\psi\pi^\pm$ decays. By fitting the above mentioned MC samples simultaneously, we constrain the fit parameters of the corresponding fit components. This results in a "MC assisted" determination of the background shapes that automatically accounts for

the statistical uncertainties of the MC. For the signal, the MC is used to determine the signal shape, but two free parameters, one for the mass scale and the other for the mass resolution, are extracted from data to accommodate for the possible data-MC differences in the signal shape.

The signal shape is parameterised as Johnson S_U [145] [5] plus Gaussian for N_3 and N_2 trigger categories, while a single Johnson S_U is used for the N_1 category. The radiative contribution of the signal (when the B radiates a γ) is also considered separately as the mass shape results skewed on the left. A Johnson S_U plus Gaussian is used for N_2 and N_3 trigger categories, while a single Johnson S_U is used for the N_1 category. The relative abundances of the two signal components (non-radiative and radiative) are extracted from the fit together with the total number of signal events.

The $J/\psi\pi^{\pm}$ events are also parameterised as a Johnson S_U plus Gaussian for all the trigger categories. All the shape parameters are determined from the MC sample.

The partially reconstructed B decays (PRD) are subdivided into three categories with slightly different shapes in the low mass region. They are parameterised with a Fermi-Dirac [6] plus an exponential or a single exponential plus a constant term, and their shape parameters are determined by three separate MC samples. The relative abundances of the three components are extracted from the fit together with the total number of PRD events.

Finally, the combinatorial background is modelled with an exponential function with the shape parameter that is left floating in the fit to be extracted from data.

Figure 8.21 shows the results of the fit for the three trigger categories. The projections on each fit variable are shown for data.

Category	$\boldsymbol{N_{J/\psi\, K^{\pm}}}$
N_1	$2257 \pm 72\,(\text{stat}) \pm 36\,(\text{syst})$
N_2	$5263 \pm 98\,(\text{stat}) \pm 74\,(\text{syst})$
N_3	$48170 \pm 355\,(\text{stat}) \pm 674\,(\text{syst})$

Table 8.8. *Signal $B^{\pm}$ candidates for each trigger category extracted from the fit on 2012 data reconstructed as $B^{\pm} \to J/\psi K^{\pm}$. Uncertainties are statistical and systematic, respectively.*

.

[5] The Johnson S_U p.d.f. is parameterised as:

$$\text{Johnson } S_U = \frac{1}{\sqrt{1+t^2}} * [t + \sqrt{1+t^2}]^{-\gamma - \frac{1}{2} * \delta * Log(t+\sqrt{1+t^2})}, \text{ where } t = \frac{m-\xi}{\lambda}.$$

[6] The Fermi-Dirac distribution is parameterised as:

$$\text{M}_{\text{FD}}(m|\mu_{FD}, \alpha_{FD}) = \frac{1}{1 + e^{\frac{m-\mu_{FD}}{\alpha_{FD}}}}$$

where α_{FD} accounts for the slope, and μ_{FD} accounts for the mass scale at which the step-like effect occurs.

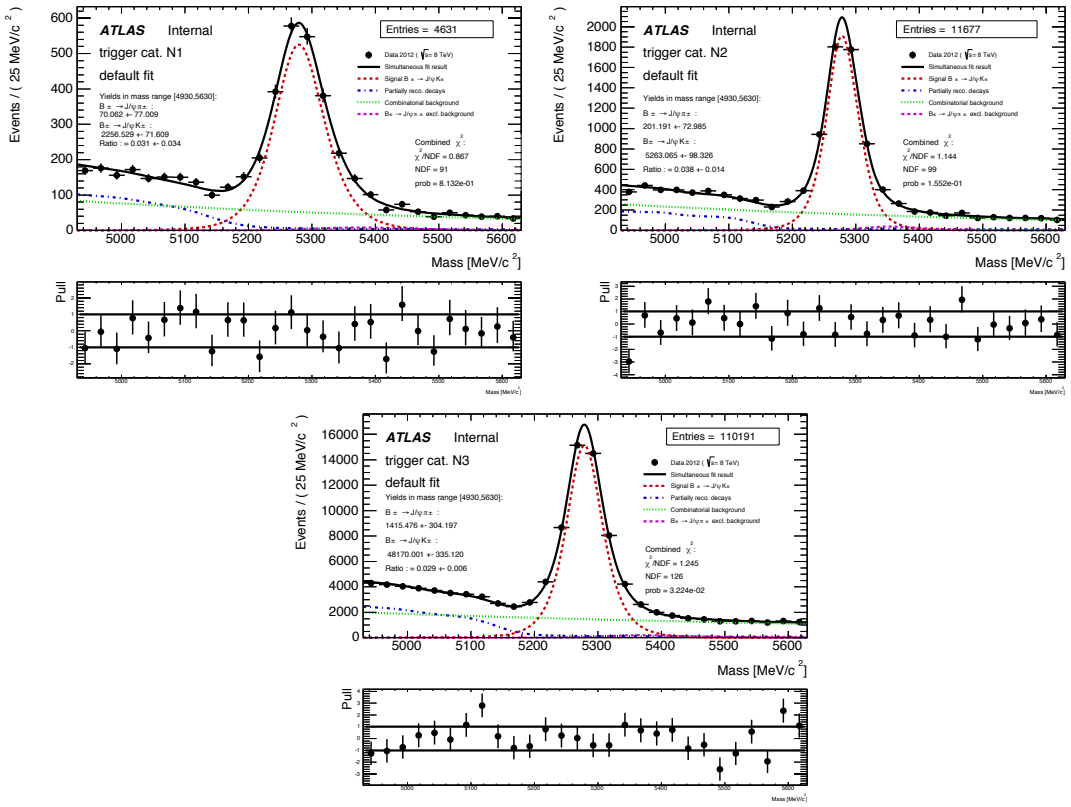


Fig. 8.21. *Fit projection on data for N_1 (top-left), N_2 (top-right) and N_3 (bottom) trigger categories. The red dashed line represents the $J/\psi K^\pm$ signal (including both radiative and non-radiative components), while the magenta dashed line represents the $J/\psi\pi^\pm$ peaking component. The blue dashed-dotted line shows all the three partially reconstructed contributions and the green dotted line represents the combinatorial background. The total of all functions is presented with the black solid line.*

Systematic effects related to the limited MC statistic are automatically considered in the simultaneous fit, while data-MC discrepancies in the signal scale and resolution are included in the fit model using the scaling factors mentioned above.

The additional systematic uncertainties, evaluated by varying the default fit model described above, take into account the kinematic differences between data and the MC used in the fit, the signal charge asymmetry, the PRD relative fractions and shapes, and the combinatorial shape. In each case, the difference with respect to the default fit is recorded and used as systematic error. The main systematic contributions come from the combinatorial shape, the PRD relative fractions and the signal charge asymmetry. The global systematic associated to the signal yield is about 1.4 %.

The number of candidates $J/\psi K^\pm$ extracted from the fit in the three trigger categories is summarised in table 8.8, where the related statistical and systematic uncertainties are also shown.

8.9 Evaluation of the Acceptances and Efficiencies

For each trigger category k, the ratio of the acceptance A and selection efficiency ε

$$R^k_{A\varepsilon} = \frac{(A \times \varepsilon)^k_{\mu^+\mu^-}}{(A \times \varepsilon)^k_{J/\psi K^\pm}} \tag{8.3}$$

is evaluated using $B^0_s \to \mu^+\mu^-$ and $B^\pm \to J/\psi K^\pm$ signal MC, after having applied to both samples the GL and DD corrections.

In the evaluation of $A \times \varepsilon$ for both signal and reference channel, the events passing the baseline and the additional selections are normalised to the events generated in the fiducial phase-space volume with $p^B_T > 8.0$ GeV and $\left|\eta^B\right| < 2.5$. As the master formula 8.1 has been modified to take into account possible differences in the three trigger categories, the ratios of the $A \times \varepsilon$ values for the signal and the reference channel are now calculated separately in each trigger category and are summarised in table 8.9.

Trigger category	$\boldsymbol{R_{A\varepsilon}}$
N_1	$0.3850 \pm 3.01\,\%$ (stat) $\pm 10.92\,\%$ (syst)
N_2	$0.3214 \pm 2.11\,\%$ (stat) $\pm 8.50\,\%$ (syst)
N_3	$0.1930 \pm 0.65\,\%$ (stat) $\pm 1.25\,\%$ (syst)
Inclusive	$0.2020 \pm 0.62\,\%$ (stat) $\pm 1.38\,\%$ (syst)

Table 8.9. *$R_{A\varepsilon}$ factors for the three triggers categories and for their logical-or. The luminosity re-weighting has been applied in the computation.*

Whereas the statistical uncertainty on the ratios is dominated by the finite statistic of the signal MC sample, the systematic uncertainty has two main components as for the 2011 analysis. The first one originates from the uncertainties of the GL and DD corrections. For the estimation of this effect a toy MC study has been performed by varying the corrections within their statistical uncertainties and re-evaluating the ratio $R_{A\varepsilon}$. The second source arises from the data-MC discrepancies in the distributions of the 22 discriminating variables. The systematic has been assessed evaluating the variation of the selection efficiency after having re-weighted the events by the ratios of the observed data and MC shapes (the procedure is the same used for the 2011 analysis detailed in section 5.8). The systematic error quoted in the table is the square root of the quadratic sum of these two components.

8.10 Studies for the Signal Yield Extraction

This section details the final configuration adopted for the invariant mass fit on data collected during Run 1. For the reasons reported in the previous chapter, the selected fit strategy is using different categories (in terms of BDT output) that are

fitted simultaneously to extract the B_s^0 yield by means of an un-binned maximum likelihood fit.

The model for describing signal and background is based on MC and on data collected in the sidebands of the search region. The expected sensitivity, possible fit bias and systematic uncertainties are evaluated using toy MC experiments.

8.10.1 $B_{s,d} \to \mu^+\mu^-$ Signal Modelling

As already discussed in section 7.2.4.5, to improve the fit sensitivity, the selected events have been classified in intervals of the continuum-BDT output. The three following bins (labelled as *bin 1*, *bin 2* and *bin 3*) have been chosen: $[0.252, 0.358]$, $[0.358, 0.458]$ and $[0.458, 1.00]$ (note that no B_s^0 events are found in the MC for continuum-BDT > 0.800). Each of the intervals corresponds to a relative signal efficiency (normalised to the number of events after the baseline, additional and fake-BDT selection) equal to 18 %, and they are ordered according to increasing *signal-to-noise* ratio. The BDT distribution for B_s^0 events is shown in figure 8.22, left side. A fourth interval labelled as *bin 0* ($0.141 <$ BDT < 0.252) has been considered, but it is found to be completely dominated by the combinatorial background that has a mass slope different with respect to the other three bins and therefore it does not contribute in a significant way to the signal extraction (see section 8.10.4 for more details).

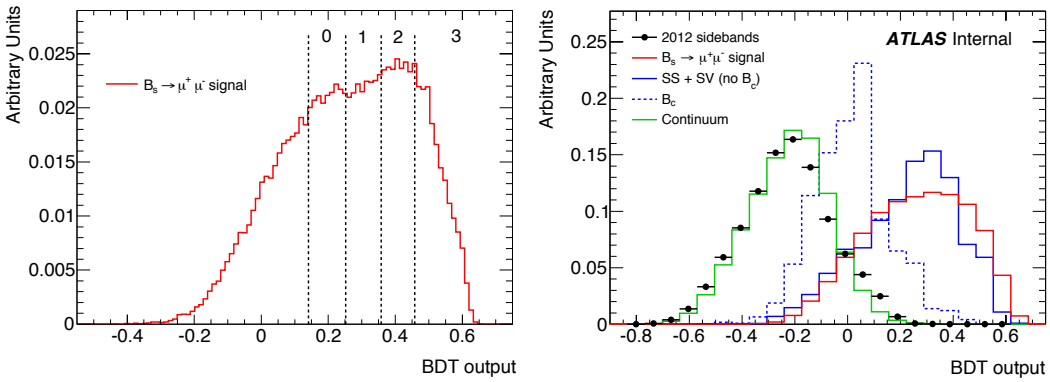


Fig. 8.22. *Left: continuum BDT distribution for signal MC. The vertical lines show the limits of the intervals used for the four-bin definition. The first bin is not used for the signal extraction. Right: continuum BDT distribution for all the components entering in the fit: signal B_s^0, partially reconstructed B events (SS-SV, see section 8.10.2), B_c decays and the continuum. The solid histograms are obtained from MC, while the black dots represent the sideband data. All distributions are normalised to unity.*

The number of B_s^0 events expected in the fiducial phase space, has been estimated to correspond to 162 events in 20 fb^{-1} of integrated luminosity collected at 8 TeV of c.m. energy[7]. The combination of the baseline and the additional selection, including

[7]The number of events has been estimated using exactly the same procedure adopted in section

the multivariate selection against fake muons, has an efficiency of 34.7 %, while the selection on the continuum-BDT has a 54 % relative efficiency in the sum of the three bins described above (continuum-BDT > 0.252). The efficiency of the whole selection chain is therefore equal to 18.7 % giving 30 B_s^0 events in total (10 in each BDT interval) generated in the toy MC experiments.

Figure 8.23 shows the B meson p_T^B and η^B and the continuum-BDT distributions for the B_s^0 and the B^0 MC simulations, normalised to unity. The reconstruction efficiencies (after the baseline, additional and fake-BDT selection) for B_s^0 (34.7 % $\pm$ 0.2 %) and B^0 (34.5 % $\pm$ 0.2 %) are well compatible within the errors. Therefore, the ratio of 11.3 % (see section 1.6) between the B_s^0 and B^0 event yield predicted by the SM can be safely applied, expecting to observe one B^0 decay every 8.9 B_s^0 events.

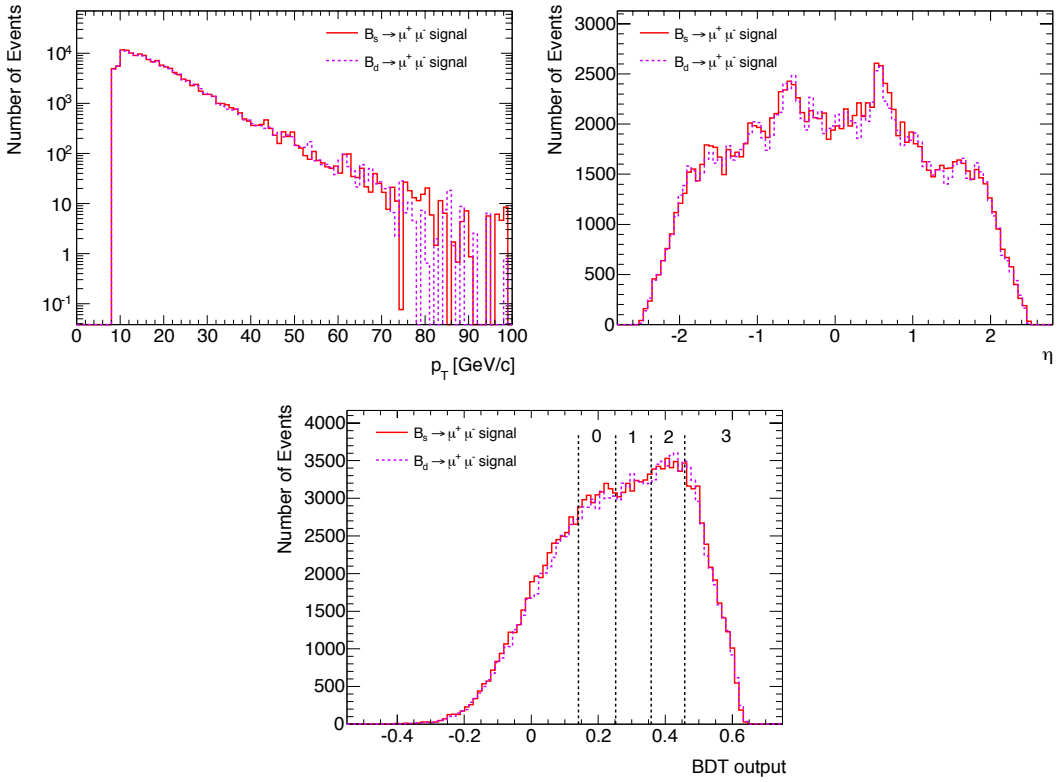


Fig. 8.23. *Distributions of the B meson p_T (top left) and η (top right), as well as of the continuum-BDT (bottom) for B_s^0 and B^0 MC simulations.*

The mass shape of the $B_s^0 \to \mu^+\mu^-$ signal is described, both in the generation and in the likelihood fit, by a superposition of two Gaussian distributions, both centred at the B_s^0 mass value (from PDG [147]). The parameters of this distribution have been

7.2.4 where the 1.7 events expected in 2011 (from equation 7.13) were rescaled considering the 2011 selection efficiency, the 2011/2012 integrated luminosity ratio and the increasing in the cross-section due to the higher c.m. energy.

extracted from MC, as shown in figure 8.24. The signal mass p.d.f. is found to be independent of the output of the continuum-BDT used to reduce the combinatorial background.

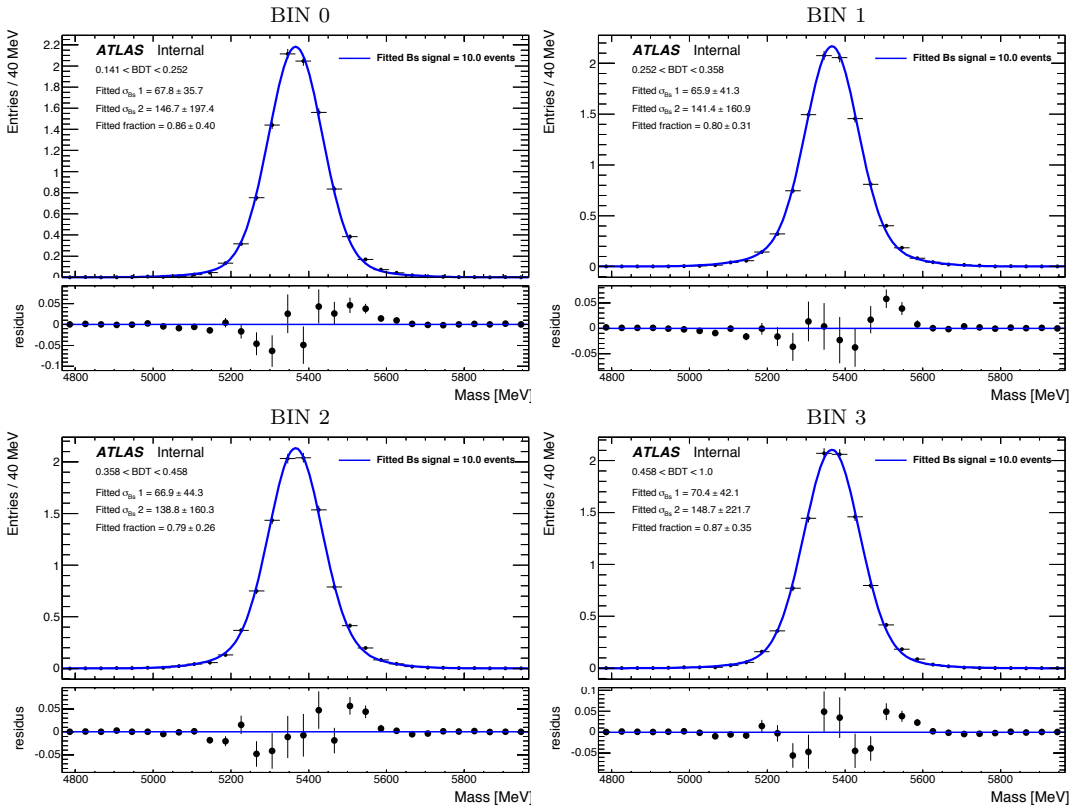


Fig. 8.24. *From left to right, from top to bottom: invariant mass distributions for the $B_s^0 \to \mu^+\mu^-$ signal from MC, in the four BDT bins, superimposed to the fit with a double-Gaussian distribution.*

Preliminary studies have been performed considering possible alternatives to this baseline description. Distributions with wider tails are needed in order to deal with non-Gaussian response of the spectrometer, with the dependence of the momentum resolution on the muon rapidity $|\eta|$ and, to a lesser extent, on the muon transverse momentum. Two models, based on the studies reported in section 8.4, have been considered:

- a (single) Gaussian distribution combined (as conditional p.d.f.) with a per-event mass error;
- a (single) Gaussian distribution, with the events binned according to the largest muon rapidity ($\eta_{\max}$) and the mass resolution depending on the rapidity bin. The resolution is uniform and better in the ATLAS barrel region, and degrades in the transition and end-cap regions.

Both variations have been discarded since they increase the complexity of the model (as the η dependence has to be inferred also in the background) and they have been found from toy experiments not to increase the significance of the signal extracted.

On figure 8.25 left, the different components of the B_s^0 mass distribution are shown, i.e. two superimposed Gaussians for the signal, and an off-set Gaussian for the radiative tail. This radiative tail is not included in the signal default p.d.f.. As it will be discussed below, tests were performed to check that those events feed into the background components of the fit and that no bias is introduced in the signal yield neglecting its presence.

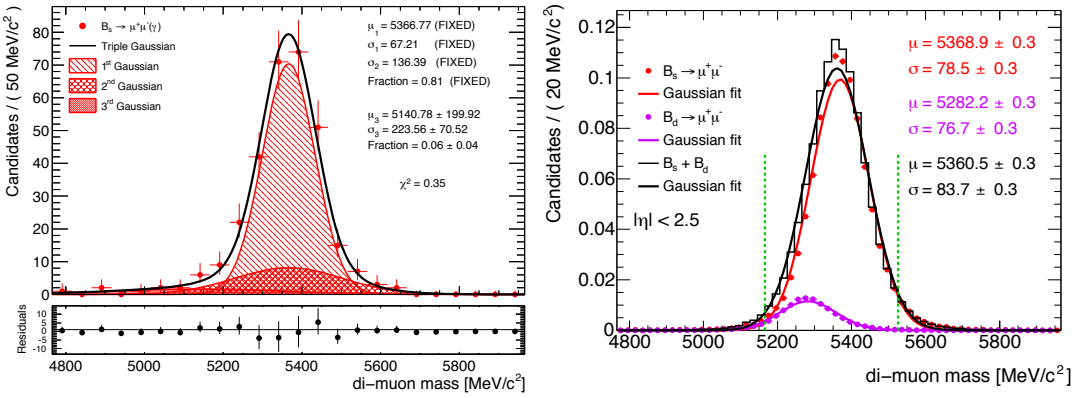


Fig. 8.25. *Left: fit to the signal including the radiative component $B_s^0 \to \mu^+\mu^-\gamma$. The first and second Gaussian are the same as in the exclusive fit, while the third takes care of the radiative tail. Right: fits to the signal mass distribution in the full pseudo-rapidity range for both the B_s^0 and the B^0 signals normalised to the relative (SM) production cross section ratio, as well as fit on the sum of these two distributions.*

The $B^0 \to \mu^+\mu^-$ mass distribution is described in the toy generation similarly to the B_s^0 one. The mass separation between the two components ($\sim$ 87 MeV) is comparable to the mass resolution ($\sim$ 80 MeV in the full η range). Figure 8.26 shows the mass distributions normalised to unity for B^0 and B_s^0 obtained from MC in the full pseudo-rapidity range as well as in the three resolution zones defined in section 8.4. Figure 8.25 right shows the mass distribution for both the B_s^0 and the B^0 signals normalised to the SM relative contribution in the full pseudo-rapidity range, as well as the fit on the sum of these two distributions. This plot underlines how the separation of the two contributions is not trivial with the given resolution (see section 8.10.7).

8.10.2 Background Components

The large sample of MC events obtained with the four-corners production has been used to study the characteristics of the different sources of background. As discussed in chapters 6 and 7, both sidebands are dominated by the combinatorial background, but in the low-mass region contributions from partially reconstructed decays are also

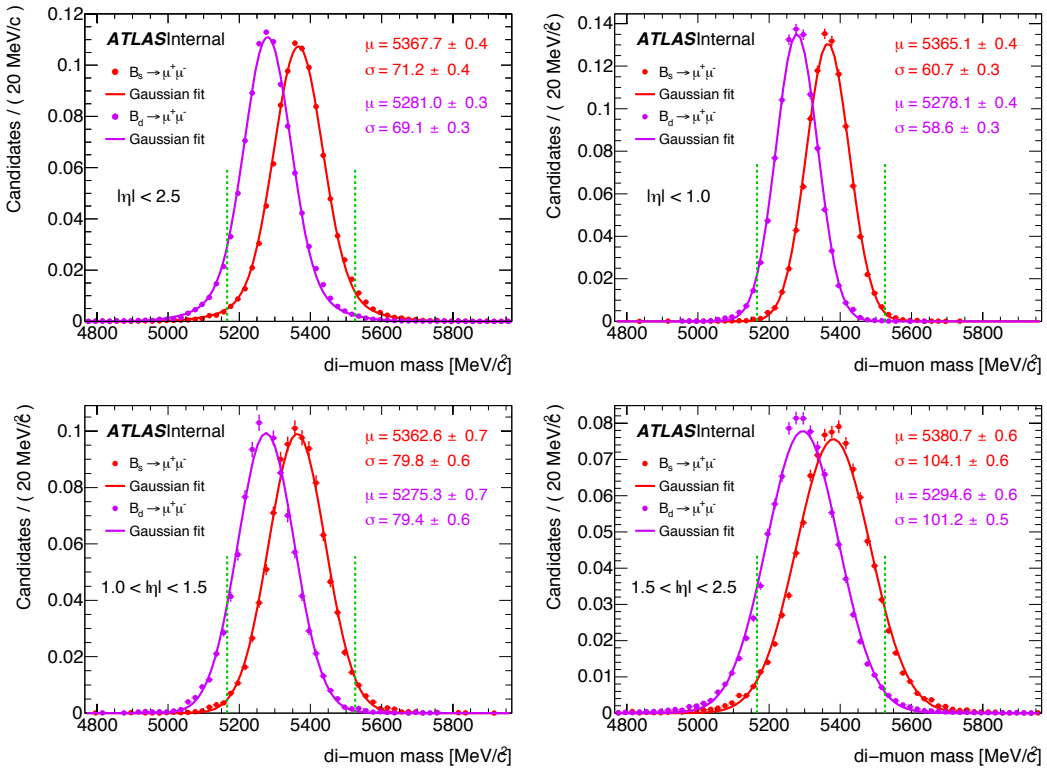


Fig. 8.26. *Fits to the mass distribution for both the B_s^0 and the B^0 signals in the full pseudo-rapidity range, as well as in the three resolution zones. The two distributions are normalised to unity, while according to the SM prediction for the branching fractions, together with the known ratio of production cross section, the signal for B^0 would be about 9 times smaller than the B_s^0 one. The fits in the different resolution zones are done with a simple Gaussian, while two Gaussians are used in the full pseudo-rapidity range (in this case the reported width σ is obtained with a simple Gaussian fit in a mass window of 150 MeV around the mass central value).*

present.

Using MC truth information, the events are classified according to their origin into the following classes:

- combinatorial background, due to opposite charged muons not from the same B_s and B^0 decays;
- same-vertex (SV) background, due to partially reconstructed B_s^0 and B^0 events containing a muon pair, such as $B^0 \to K\,\mu^+\mu^-$;
- same-side (SS) background, due to oppositely charged muons from the same b quark: $b \to c\,\mu^-\nu \to s(d)\,\mu^+\mu^-\nu$;
- B_c background, for example $B_c \to J/\psi\,\mu^+\nu \to \mu^+\mu^-\mu^+\nu$.

Figure 8.27 shows the invariant mass distribution for SV background (left side) as well

as SS and B_c decays (right side) with the relative composition of the different decays.

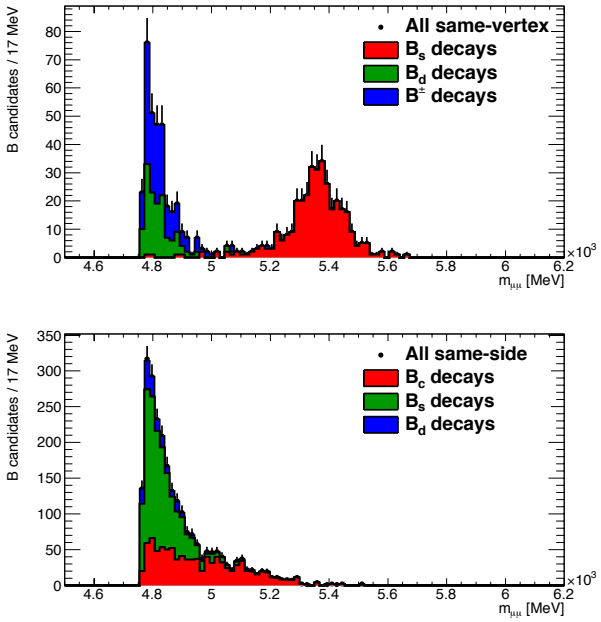


Fig. 8.27. *Invariant mass distribution for SV (top) and SS (bottom) events.*

Figure 8.28 shows the comparison for the invariant mass and the continuum-BDT output distributions of the sideband data with the background MC distributions described above. The following features may be remarked:

- Figure 8.28, left: the combinatorial di-muon mass background has a mild linear shape which is confirmed by the data. The other backgrounds accumulate at the low-mass values.
- Figure 8.28, right: for BDT output less than 0.4, the background is dominated by the combinatorial component, while above that value, SV+SS and, to smaller extent, B_c background are selected being B events. This is shown also in figure 8.22, right-side, looking at the continuum-BDT distributions: SS+SV events have a BDT response similar to the one of signal events, while the B_c background extends over both regions populated by signal and by combinatorial events.

The combinatorial background, highly dominant before the multivariate selection, is reduced in the bins with high continuum-BDT values, and its small mass dependence facilitates its interpolation from the sidebands into the signal region. The SS+SV events behave like the signal in terms of BDT output and they are enhanced at high continuum-BDT values, but accumulating in the low-mass region and therefore well separated in mass with respect to the signal. The B_c component behaves in an intermediate way between the other two, but due to its negligible cross section its presence is less relevant, as discussed below.

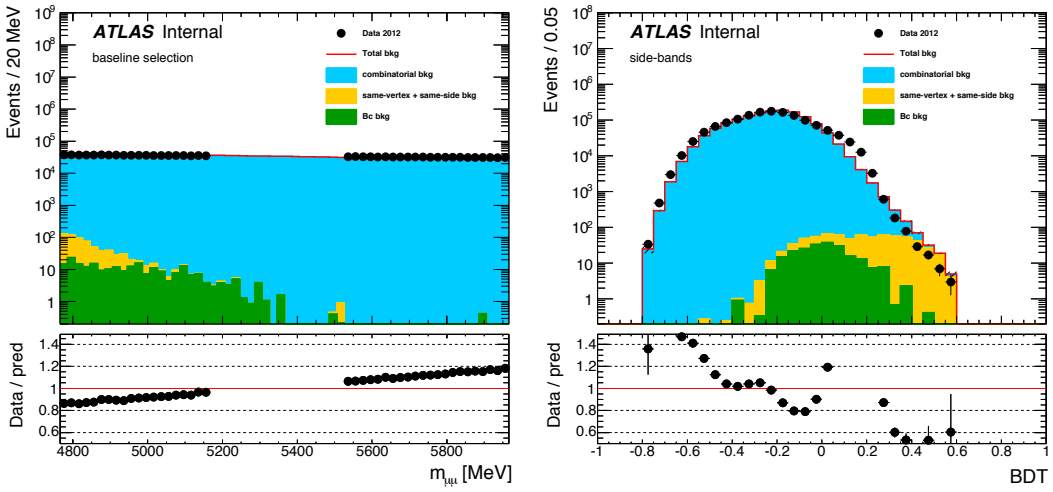


Fig. 8.28. *Distributions of the invariant mass (left) and continuum-BDT output (right) for sideband data and for the four-corners MC sample after all selection cuts, except the one in the continuum BDT variable.*

The distributions of the invariant mass, the number of reconstructed PV, the B meson p_T^B and η^B, the continuum-BDT variable, and the discriminating variables α_{2D}, L_{xy} and ΔR as obtained in MC and in sideband data after the lower cut on the continuum-BDT (continuum-BDT > 0.252, three bins together) are shown in figures 8.29 and 8.30.

As already discussed in section 8.6, the peaking background is composed of $B \to hh'$, mainly $B_s^0 \to K^+K^-$ and $B^0 \to K^\pm\pi^\mp$, in which both hadrons are misidentified as muons. Due to the mass distortion related to the $K \to \mu$ mass assignment, and the smaller one for $\pi \to \mu$, the mass distribution of these events is substantially superimposed with the B^0 signal, as shown in figure 8.16. Thanks to the specific multi-variate selection for fake muons, less than 1 peaking background event is expected in the sum of the three BDT bins, corresponding to about 2.5 % of the B_s^0 signal and about 20 % of the expected B^0 event yield.

8.10.3 Background Modelling

The **combinatorial** (opposite-side) di-muons background mass shape is described with a Chebychev first order polynomial $f(x) = 1+\alpha\, T_1(x) = 1+\alpha\, x/1200$ MeV. Figure 8.31 shows the mass distribution of the events belonging to this class of background in the four continuum-BDT bins, as described by MC (the un-binned linear fit is superimposed). The slope of the distribution is similar in the second and third continuum-BDT intervals, and compatible, within the large statistical error, with the fourth interval. The first interval is shown for comparison but not used for signal extraction having a slope slightly different with respect to the other three (see section

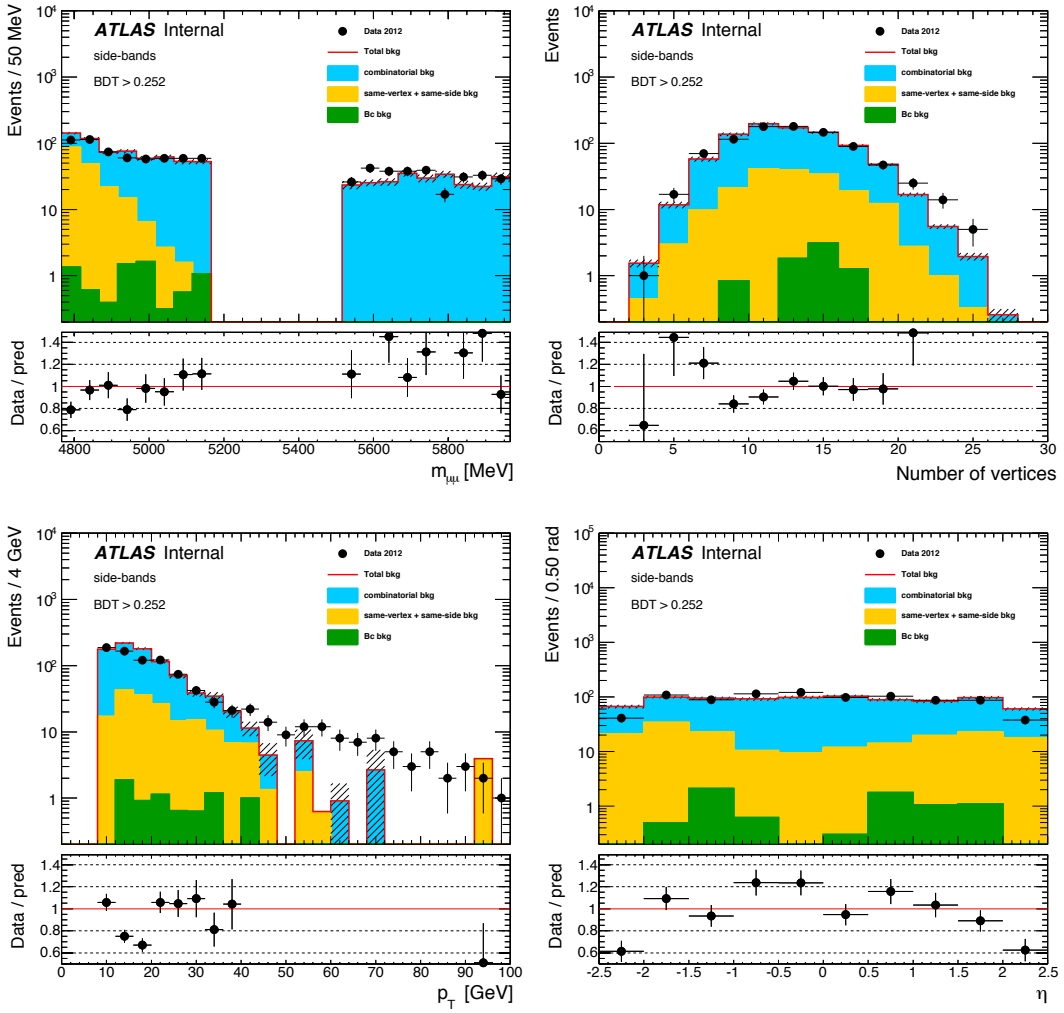


Fig. 8.29. *Distributions of sideband data and four-corners MC sample after all selections, and after requiring continuum-BDT > 0.252. The normalisation of the four-corners sample is done after the selection on continuum-BDT. From left to right, from top to bottom: invariant mass, number of primary vertices, B meson p_T and η.*

8.10.4 for more details).

In the **same-side** and **same-vertex (SS-SV)** background, the mass distribution of the two muons is peaked far below the signal region, and we are sensitive to a tail of the distribution determined by kinematic limits, detector resolution effects and missing particles in the final state. Figure 8.32 shows the exponential p.d.f. $f(x) = \exp(\alpha\, x)$ used to fit the shape for MC events in this class. The shapes of the distributions are compatible (within the errors) across the four bins in line with the negligible correlation between mass and continuum-BDT in B events.

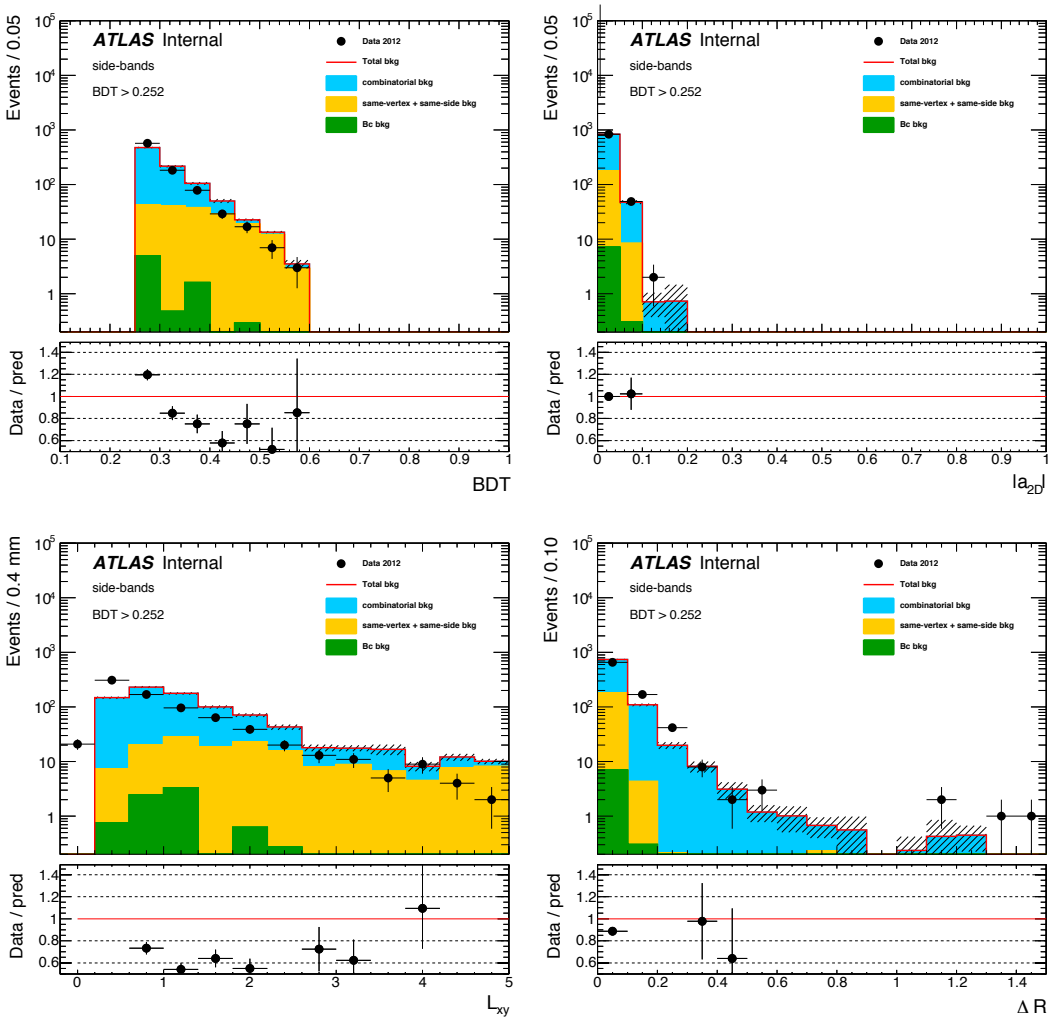


Fig. 8.30. *Distributions of sideband data and four-corners MC sample after all selections, and after requiring the continuum-BDT > 0.252. The normalisation of the four-corners sample is done after the selection on continuum-BDT. From left to right, from top to bottom: continuum-BDT variable, α_{2D}, L_{xy} and ΔR.*

The $\boldsymbol{B_c}$ component of the background is mainly due to a small number of events in which B_c decays into $J/\psi\,\mu\,\nu$. The mass shape is smoothly decreasing towards the signal region and it can be fitted with an exponential function ($f(x) = \exp(\alpha\, x)$). Figure 8.33 shows the fit to the B_c events in the invariant mass distribution in each continuum-BDT bin. Given the relatively smaller amplitude, these events are expected to feed into the continuum and partially in the SS-SV events. Therefore no p.d.f. is added in the fit to model the B_c component of the background as further discussed below. This assumption has been checked using toy MC experiments including the B_c

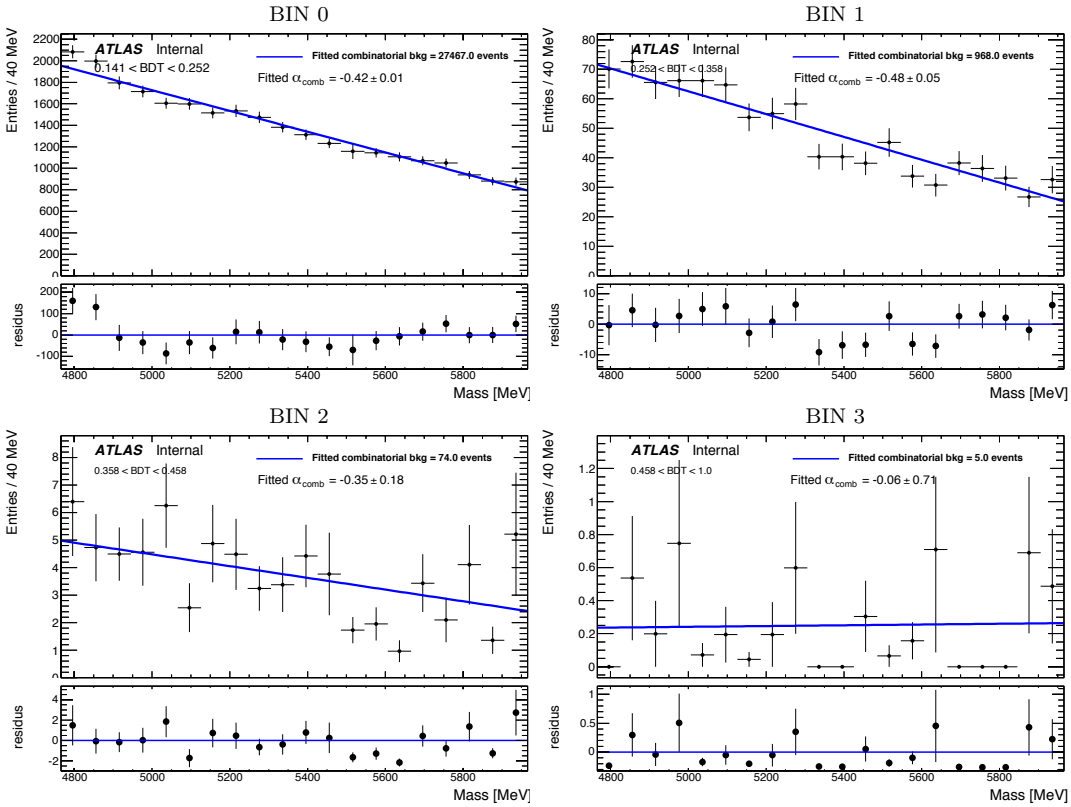


Fig. 8.31. *From left to right, from top to bottom: invariant mass distributions for continuum events from the four-corners MC sample normalised to the data in the sidebands. Each figure corresponds to events in a specific continuum-BDT bin. The number of events over the full mass range and the fit result are reported in the legends and in table 8.10.*

component in the generation, but not in the fit model, and no bias in the yield and shape parameters is found.

The **semi-leptonic** background is due to three-body semi-leptonic B decays feeding into our final selection through a misidentification $h \to \mu$ ($h = \pi, K, p$), in the limit of low energy neutrinos. In particular, $B^0 \to \pi\,\mu\,\nu$ and $B_s^0 \to K\,\mu\,\nu$ are expected to be dominant in this kind of background, together with $\Lambda_b \to p\,\mu\,\nu$. The mass distribution of these events is shown in figure 8.34. The B^0 and B_s^0 components have a very similar shape, fitted with the tails of Gaussian distributions. The shape of the Λ_b distribution, that extends closer to the signal region, is also shown, but its relative normalisation is found to be negligible because of a very low probability of misidentifying the proton as muon in ATLAS (see section 8.6). As all the B-like backgrounds, the invariant mass distribution of the semi-leptonic decays has a very small dependence on the continuum-BDT value. We expect in total $1.6, 1.7, 1.7, 1.7$ events in the four continuum-BDT bins for the full mass and η range. The semi-leptonic background mass shape is expected to be fitted by the first-order polynomial and the exponential p.d.f.s, and it is not

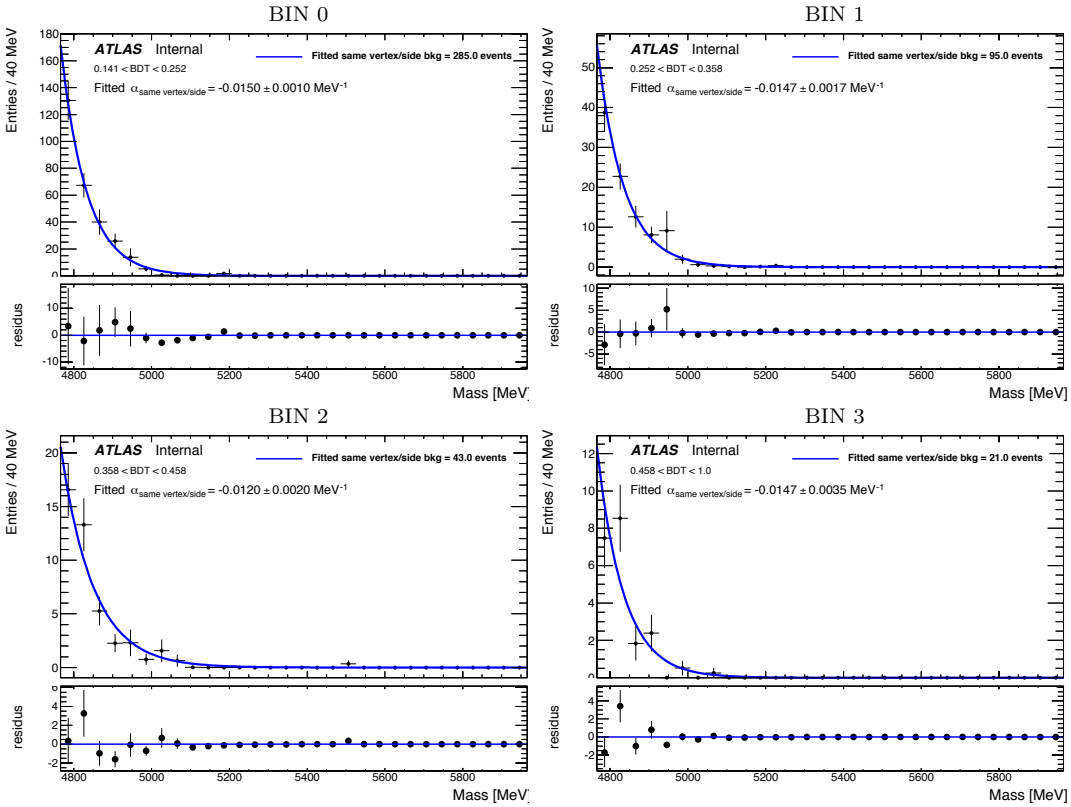


Fig. 8.32. *From left to right, from top to bottom: invariant mass distributions for same-side (SS) and same-vertex (SV) events from the four-corners MC sample normalised to the data in the sidebands. Each figure corresponds to events in a specific continuum-BDT bin. The number of events and the fit result are reported in the legends and in table 8.10.*

included in the fit model as further discussed below. This assumption has been checked with toy MC experiments using the same procedure adopted for the B_c component, and no bias in the results is found.

8.10.4 Fit of the Background Components to Sideband Data

In each bin in the continuum-BDT, the mass distribution of the sideband data is fitted to check the parameterisation of the shape extracted from the MC, and interpolated in the signal region to estimate the expected number of events.

Figures 8.31, 8.32, 8.33 and 8.34 show what was already discussed in the previous sections: for the main type of background events the mass distribution of the muon pairs is equal across the continuum-BDT bins or compatible within the statistical uncertainties. This condition is particularly useful for the continuum component as it can be applied as a constraint in the fit.

This is verified by considering the correlation between the continuum-BDT and the

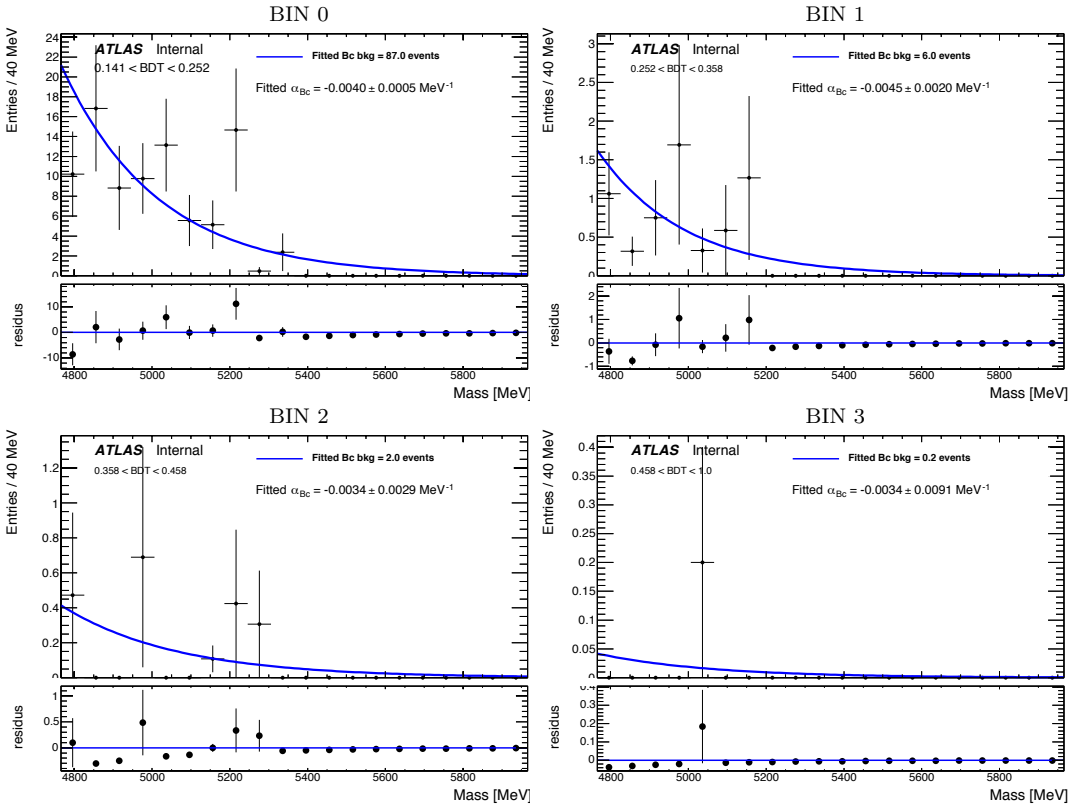


Fig. 8.33. *From left to right, from top to bottom: invariant mass distributions for B_c events from the four-corners MC sample normalised to the data in the sidebands. Each figure corresponds to events in a specific continuum-BDT bin. The number of events and the fit result are reported in the legends and in table 8.10.*

di-muon mass: while in the signal and in the signal-like events we found no correlation, in the continuum background we found a small correlation between the mass shape and the BDT value.

Top plots in figure 8.35 show the di-muon invariant mass distributions of combinatorial MC events and sideband data in various continuum-BDT bins: the black points correspond to all the events cut away from our continuum-BDT selection (continuum-BDT < 0.141), while the red and blue points represent the first and second bin respectively. The green points correspond to the merging of the last two bins as they have a reduced statistics. Bottom plots in figure 8.35 contain the slope values coming from the fit to the continuum background MC and sideband data mass shape as function of continuum-BDT bins covering the whole BDT range. There is a correlation effect that is reduced once the continuum-BDT range is limited to specific values, like for example to the high (signal-like) ones. The residual correlation is considered as systematic uncertainty of the fit and it is estimated leaving the combinatorial background shape floating with a Gaussian constraint (see section 8.10.8). The choice

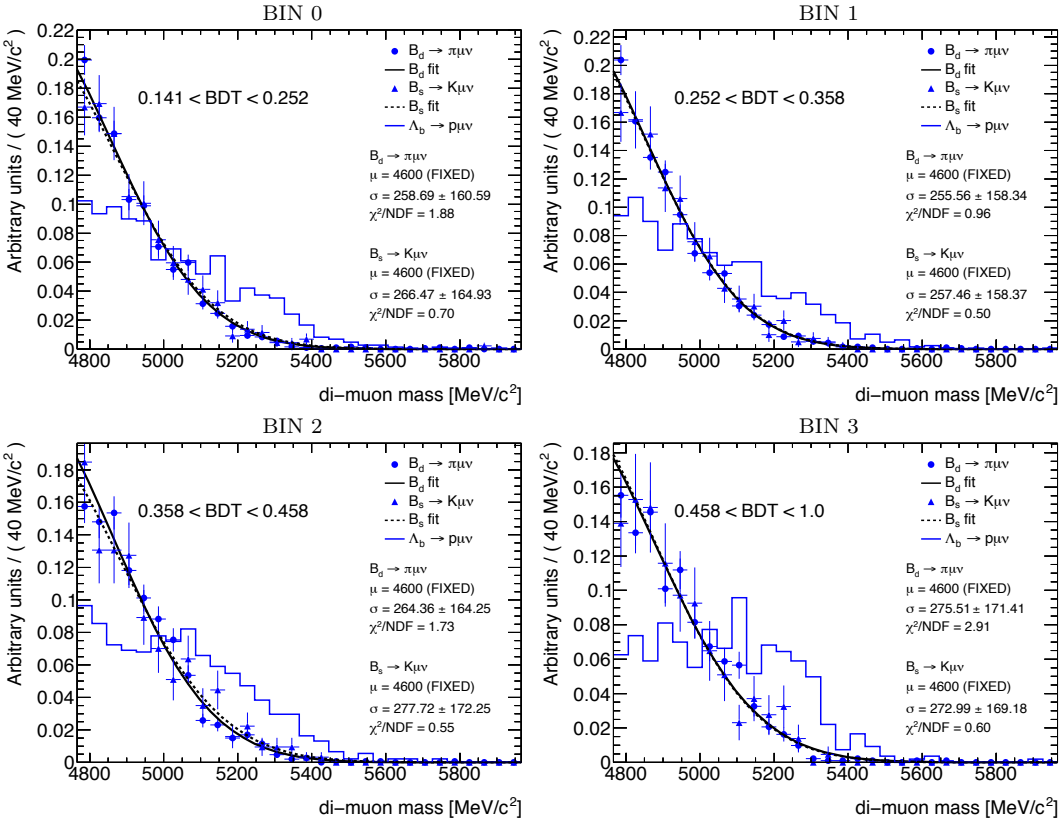


Fig. 8.34. *From left to right, from top to bottom: invariant mass distributions for semi-leptonic events from the exclusive MC samples normalised to unity. Each figure corresponds to events in a specific continuum BDT bin. The B_s and B^0 distributions are fitted using the tails of Gaussian distributions and the results are reported in the plots.*

of the number of continuum-BDT bins to be used in the fit is driven by this dependence: clearly the first bin (red points) has a shape marginally compatible with the other used in the fit. Moreover, as the data-MC comparison for this shape is not perfect, it is much safer to stay away from the step-like structure appearing after continuum-BDT values of -0.10 in figure 8.35. The first bin can not be used to constrain the continuum shape in the fit, making it essentially non influential on the final result.

If the slope constraint across the bins is not applied the values of the shape parameters would be affected by large uncertainties in some bins. In particular, the mass dependence of the combinatorial background would not be reliably determined in the last bin.

The shapes of the combinatorial and SS+SV background components are constrained to be equal in the three continuum-BDT bins and the values of the shape parameters are left free in the fit to be determined on data.

The fitting procedure has been tested with toy MC experiments. The toys are initially generated including all background contributions described above. However,

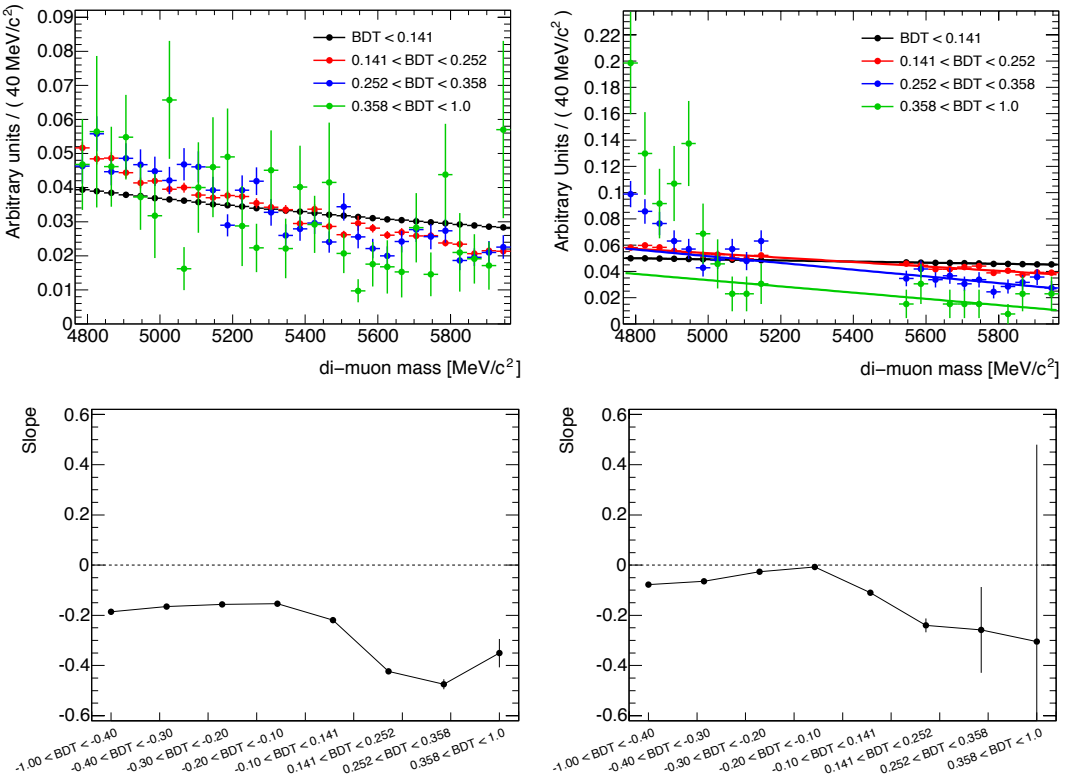


Fig. 8.35. *Top: invariant mass distributions as function of continuum-BDT ranges in the four-corners background MC events (left) and in the sideband events superimposed with the linear fits (right). Bottom: values of the mass shape slopes as function of continuum-BDT bins for the four-corners continuum MC fits (left) and for the sideband fits (right).*

as already discussed in the previous section, tests performed have shown that only the two larger background components, namely combinatorial and SV+SS should be considered in the fit, while other classes of background, B_c and semi-leptonic decays, have small amplitudes and feed in the continuum or SS+SV components. Therefore only the two former background components are included in the fit. No additional bias on the extracted signal yield has been observed with this fit configuration even generating all background contributions.

The result of the fit to the sideband data is reported in table 8.10. For each continuum-BDT bin and for the two major classes of background the table shows the number of events obtained in the signal region (interpolated from the sidebands) and those obtained for the full mass range (sidebands+signal region). These numbers, and the values of the shape parameters obtained from the simultaneous fit, can be compared to the results obtained in the MC, also shown in the table. The first bin with lower BDT values is also reported in this case. Figure 8.36 reports the data in the sidebands, superimposed with the fit.

Continuum-BDT interval		**0.141–0.252**	**0.252–0.358**	**0.358–0.458**	**0.458–1.0**
Events from fit to s.b. data	Comb. bkg. SV+SS+B_c		942 ± 63 115 ± 40	46 ± 13 60 ± 11	4 ± 4 21 ± 5
Events from fit to signal region	Comb. bkg. SV+SS+B_c		286 4.5	14.2 2.4	1.3 0.8
Bkg. shape from s.b. data	α_{comb} $\alpha_{\mathrm{SS+SV+Bc}}$		-0.31 ± 0.09 -0.008 ± 0.002 MeV^{-1}		
Bkg. shape from fit to MC	α_{comb} $\alpha_{\mathrm{SS+SV}}$ [MeV^{-1}] α_{Bc} [MeV^{-1}]	-0.42 ± 0.01 -0.0151 ± 0.0010 -0.0038 ± 0.0005	-0.48 ± 0.05 -0.0147 ± 0.0017 -0.0048 ± 0.0021	-0.36 ± 0.18 -0.0121 ± 0.0020 -0.0034 ± 0.0028	$+0.16 \pm 0.69$ -0.0147 ± 0.0035 N.A.
Bkg. events in MC	Comb. bkg. SV SS B_c	27300(F) ; 8130 (SR) 52 (F) ; 0 (SR) 230 (F) ; 1 (SR) 85 (F) ; 18 (SR)	947 (F) ; 273 (SR) 22 (F) ; 0.2 (SR) 70 (F) ; 0.4 (SR) 6 (F) ; 0.2 (SR)	70 (F) ; 21 (SR) 7 (F) ; 0 (SR) 35 (F) ; 0.3 (SR) 2 (F) ; 0.7 (SR)	4 (F) ; 1.0 (SR) 3 (F) ; 0 (SR) 17 (F) ; 0 (SR) 0.2 (F) ; 0 (SR)

Table 8.10. *Top six lines: fit to the sideband data with background description based on MC, namely using a first order Chebychev polynomial for the combinatorial background and an exponential for the SV+SS background, with the shapes parameters constrained to be equal in the three BDT intervals, and the normalisations extracted from the fit over the full mass range. The number of events in the different components are given for the full mass range (4766–5966 MeV) and the signal region alone (5166–5526 MeV). Bottom seven lines: backgrounds shape parameters and composition from the four-corners MC sample, normalised to the data collected in the sideband regions in each BDT bin. The background composition is given for the full mass range (F) and the signal region (SR).*

As an illustration of what we expect to obtain after unblinding the signal region, plots in figure 8.37 show the invariant mass distributions as obtained from an example toy experiment, superimposed with the fit of the two background components and the B_s^0 signal. The toy experiment events have been generated using the background shapes and normalisations obtained from the fit to the sideband data shown in figure 8.36. One B^0 and ten B_s^0 events have been generated in each bin, using the double Gaussian p.d.f.s extracted from MC.

8.10.5 Summary of The Fit Configuration

To summarise, the baseline signal fit includes the following 3 p.d.f.s:

- signal p.d.f.: sum of 2 Gaussians centred at the B_s^0 mass. The widths of the Gaussians and their relative fraction, assumed to be identical in all continuum-BDT bins, are taken from MC and fixed in the fit.

- Continuum background p.d.f.: first order polynomial. The slope is assumed to be identical in all continuum-BDT bins and it is a free parameter of the fit.

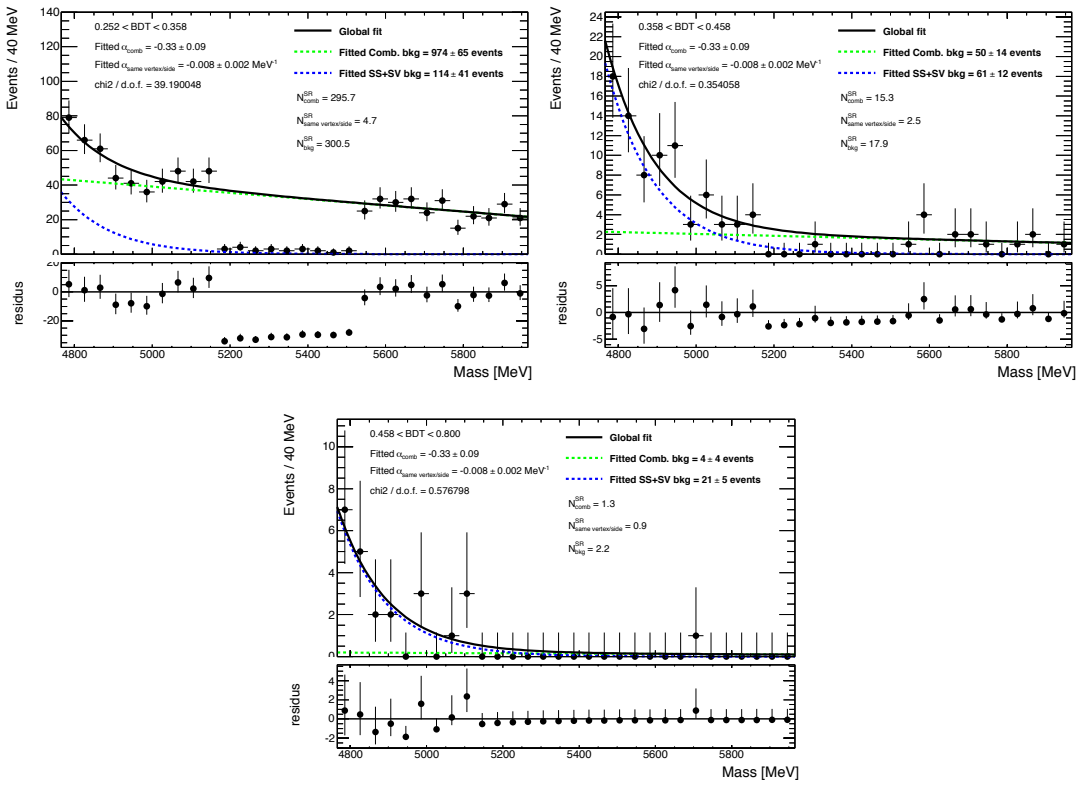


Fig. 8.36. *Invariant mass distributions in sideband events superimposed with the fit to the two main background components.*

- SS+SV background p.d.f.: exponential. The shape is assumed to be identical in all continuum-BDT bins and it is a free parameter of the fit.

The following 9 free parameters are determined by the fit:

- the total number of signal events (1 parameter), being the fraction of signal events in each continuum-BDT bin fixed at 1/3;
- the number of continuum background events in each continuum-BDT bin (3 free parameters) in addition to the slope (1 parameter);
- the number of SS+SV background events in each continuum-BDT bin (3 free parameters) in addition to the slope (1 parameter).

The events contained in the first and second bin allow to extract the shapes of the combinatorial and low-mass background, while the second and in particular the third bin provide sensitivity to the signal.

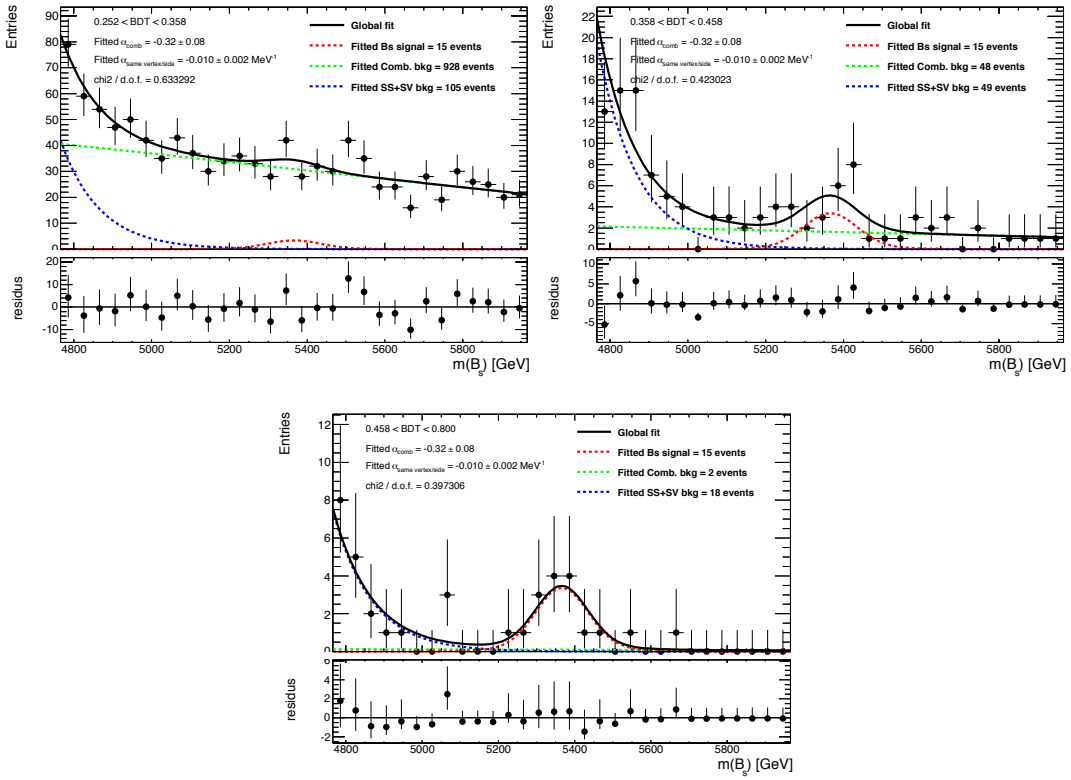


Fig. 8.37. *Invariant mass distributions as obtained from a typical toy experiment, superimposed with the fit to the two background components and the B_s^0 signal.*

8.10.6 Fit to B_s^0 Signal and Background with Toy Experiments

Besides testing the reliability of the fit procedure, toy MC experiments have been used to evaluate the expected sensitivity of the analysis. The shapes and normalisations of the two background components have been generated using the values obtained from the fit to the sideband data discussed in the previous section. Signal events have been generated following the SM predictions and the values for acceptance and efficiency discussed in section 8.10.1.

Due to the poor mass resolution (r.m.s. value similar to the difference in mass between B_s^0 and B^0), and the relatively small amount of B^0 events, no (or a rather small) sensitivity is expected on the $B^0 \to \mu^+\mu^-$ channel. The results obtained fitting the $B_s^0 \to \mu^+\mu^-$ component alone are discussed first in this section.

Toy MC experiments were seeded with 30 (3×10) B_s^0 events, 3 (3×1) B^0 events, and background parameters as shown in table 8.10. The significance of the signal is obtained from the likelihood ratio of the *signal-plus-background* over *background-only* hypothesis, and is equal to $4.1 \pm 1.0\ \sigma$. Figure 8.38 shows the results obtained. The pull distribution of the signal shows a small bias of about 1.2 events ($\sim 4\,\%$) (see

below).

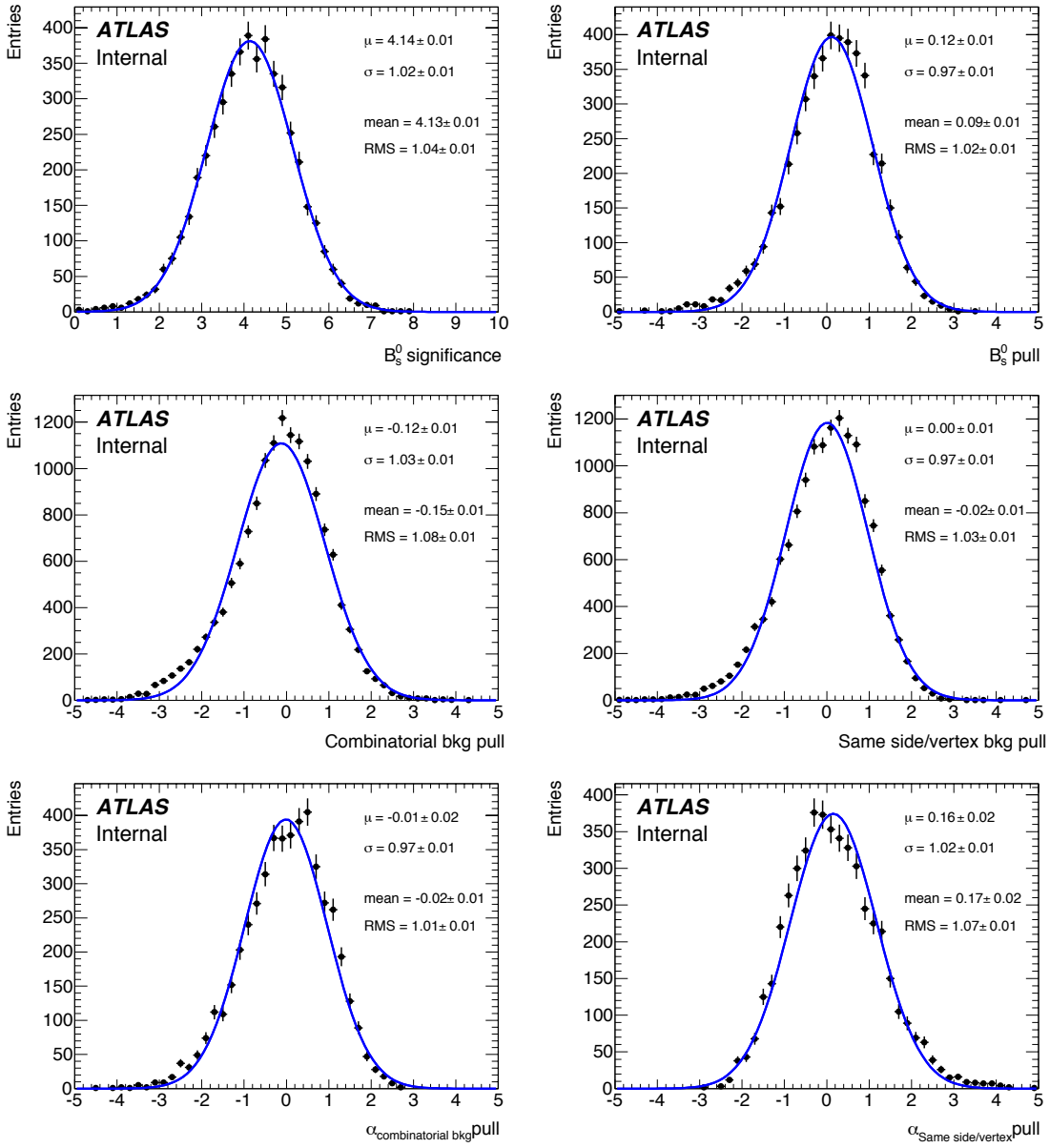


Fig. 8.38. *Fit to B_s^0 signal, as described in the text. From left to right, from top to bottom: B_s^0 signal significance, pull distributions for signal, combinatorial background, SV+SS background, and background shape parameters.*

The bias and the linearity of the fit have been tested varying the number of B^0 and adding different numbers of $B \to hh'$ resonant background used to seed the toy MC experiments, as shown in table 8.11. B^0 and $B \to hh'$ events tend to feed into the

B_s^0 signal. Even removing these two contributions a small bias in the number of B_s^0 is still present and appears to be related to small-numbers fluctuations in the current fit configuration: if the seed for the number of B_s^0 events is increased from 30 to 300, the relative bias in the observed numbers decreases from 5 % to 1 %. The results show that in the expected conditions (0.6 $B \to hh'$ and 3 B^0 events, according to SM), the bias of about 1 event is much smaller than the uncertainty of about ± 7.5 events. However, the B_s^0 yield extracted from the un-blinded data will be corrected for this bias.

Seeded number of B^0	Seeded number of $B \to hh'$	Bias on B_s Number	[%]	Significance
0	0	-1.6	-5.3	3.95 ± 1.04
3	0	+1.2	+4.3	4.13 ± 1.03
6	0	+2.8	+10.3	4.20 ± 1.06
0	0.6	-1.3	-4.5	3.97 ± 1.01
3	0.6	+0.8	+2.7	4.14 ± 1.02
6	0.6	+2.7	+9.1	4.20 ± 1.07

Table 8.11. *Bias on the extraction of the B_s^0 signal alone in toys generated with a variable number of B^0 and $B \to hh'$. The numbers reported in the table refer to the sum of the three BDT bins.*

8.10.7 Fit to B_s and B^0

Preliminary tests confirmed that the unfavourable ratio between mass resolution and mass separation between B_s^0 and B^0 limits the accuracy in determining separately the branching fractions of the two decays. With toy MC experiments seeded by SM predictions (namely: 30 B_s^0 events, 3 B^0 , and neglecting the 0.6 events contribution from peaking background), the expected significance obtained fitting separately the two contributions is 3.0 ± 1.0 σ for B_s^0 and 0.9 ± 0.7 σ for B^0. The results obtained are shown in figure 8.39.

The recent average between the LHCb and CMS experiments shows significant deviations from the SM predictions, with the B_s^0 branching ratio 1.2 σ smaller and the one for B^0 about 2.2 σ higher (see section 1.6 for more details). Including the production factor ($f_d/f_s = 3.83 \pm 0.22$), the SM prediction for the ratio of events s/d is 8.9 ± 1.1, while the corresponding LHCb/CMS number is 1.8 ± 0.8. Given the large size and the direction of the deviation between these numbers, the data from ATLAS might contribute to clarify the situation. Figure 8.40 illustrates the likelihood contours for two examples of simultaneous fit to B_s^0 and B^0, obtained when the signal is generated according to the SM prediction, or to the LHCb/CMS average.

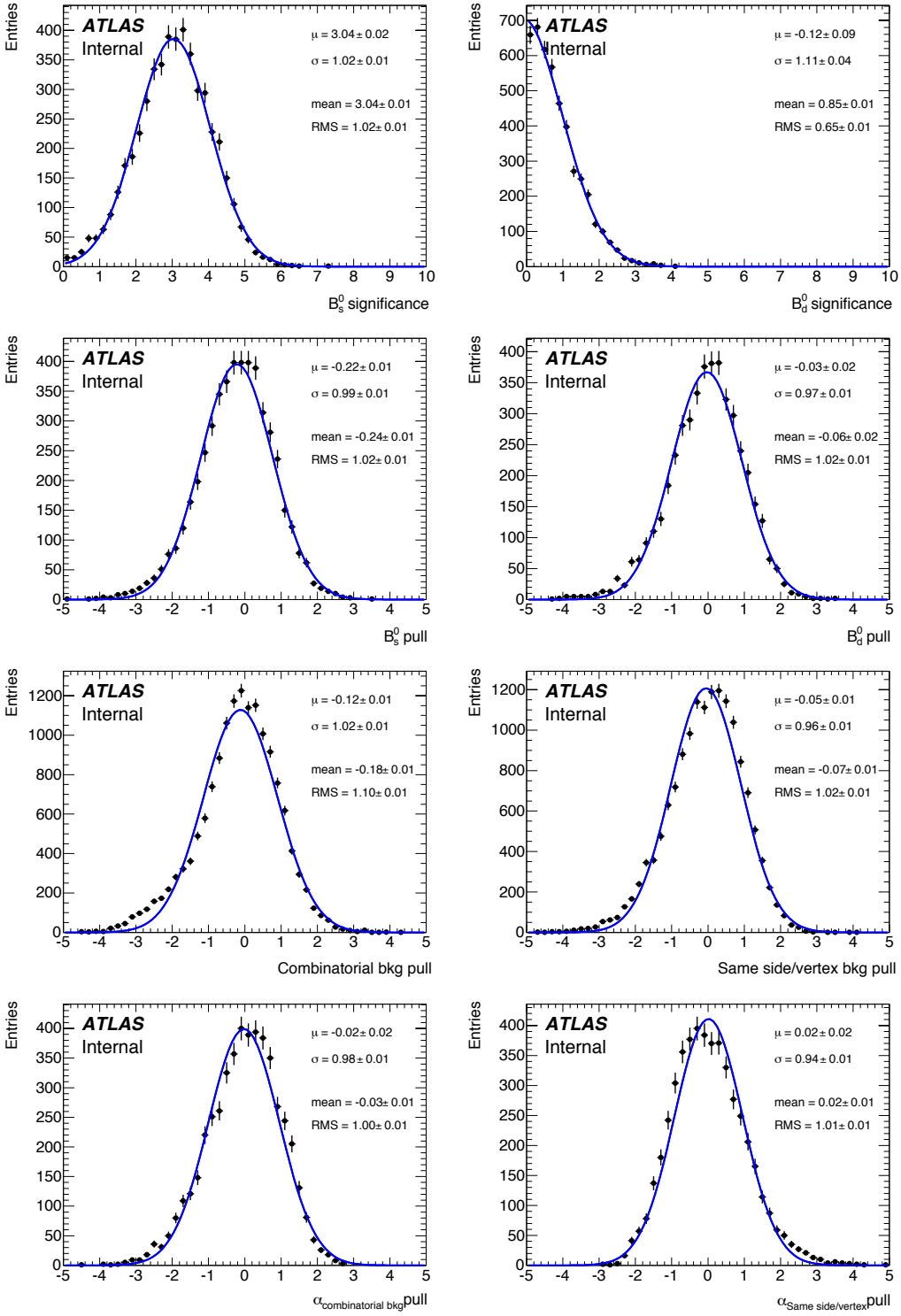


Fig. 8.39. *Simultaneous fit to B_s^0 and B^0. From left to right, from top to bottom: B_s^0 , B^0 signal significance, pull distributions for signal, combinatorial background, SV+SS background, and background shape parameters.*

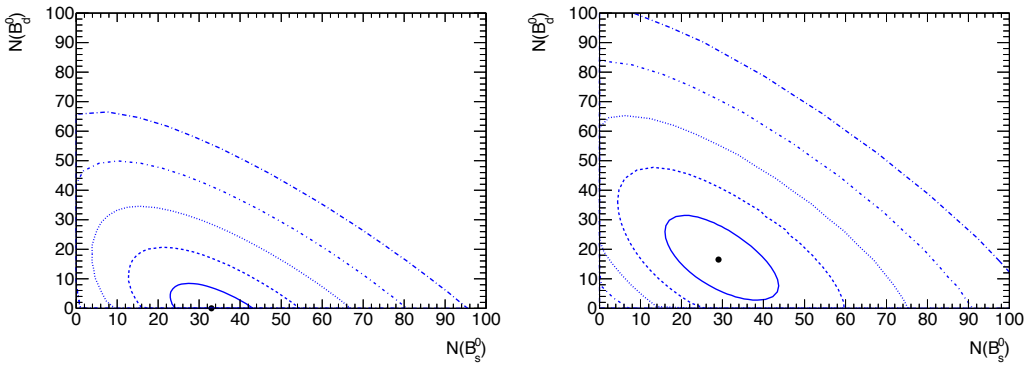


Fig. 8.40. *Examples of likelihood contours for $\Delta(2 \cdot \log \mathcal{L}) = n_\sigma^2$, for toy MC experiments with signal generated according to the SM prediction (left), and for an enhanced B^0 branching ratio (right).*

8.10.8 Systematic Uncertainties on the Signal Fit

The following variations to the baseline fit have been done to estimate the impact of the experimental and theoretical uncertainties on the B_s^0 signal yield.

- The signal acceptance has been modified through the procedure of re-weighting the distributions of the discriminating variables used in the continuum-BDT classifier with respect to the data-MC discrepancies. A negligible impact on the fitted number of signal events has been found.
- A momentum scale uncertainty of $\pm\,0.05\,\%$ has been estimated in previous B-physics ATLAS analyses and corresponds to a shift of $\pm\,2.7$ MeV on the B_s^0 (and B^0) mass. The sensitivity of the signal fit to the mass scale uncertainty has been tested by shifting the signal mean by 1 and 5 MeV, obtaining a bias in the number of extracted signal events at the level of 1 %. This small value can be explained by the tiny shift compared to the resolution of about 80 MeV.
- The sensitivity to the uncertainty on the mass resolution has been checked by generating toy MC experiments with the width of the signal p.d.f. increased or reduced by 6 % (MC mass resolution uncertainty), and fitting it with the baseline model. The difference compared to the nominal configuration was found to be negligible.
- The radiating component ($B_s^0 \to \mu\,\mu\,\gamma$), modelled from the four-corners MC, is included in the generation of toy MC experiments. This contribution is described with a Gaussian displaced by about $-\,200$ MeV, with an amplitude corresponding to about 6 % of the signal (see section 8.10.1). In this test the component is included in the generation of toy MC experiments but not in the fit to signal and background. The results obtained show how this contribution is effectively absorbed in the description of the low-mass side background and, given its size, the impact on the fit is negligible.

- The BDT intervals have been modified allowing larger (smaller) acceptance in the one with better *signal-to-noise* ratio, and no significant impact on the signal significance was observed.
- The signal acceptance in each bin has been coherently increased from 18 to 20 %, and no significant impact on the signal significance was observed.

The stability of the results has also been checked using different assumptions for the generation and the fit of the two background components.

a) Toy experiments have been generated with a different continuum background shape in each bin, randomly generated from a Gaussian with a mean and width to span the values of the shape in MC (figure 8.31) or data (figure 8.36). In the fit, the shape is assumed to be equal in all bins (*fit 5* and *fit 6* in table 8.12).

b) Toy experiments have been generated with a different low-mass background shape in each bin, randomly generated from a Gaussian with a mean and width to span the values of the shape in MC (figure 8.32) or the data (figure 8.36). In the fit, the shape is assumed to be equal in all bins (*fit 7* and *fit 8* in table 8.12).

c) Toy experiments have been generated with the same continuum and low-mass backgrounds shapes in all bins. In the fit, the shapes are assumed to be different in each bin. This is done separately for each background component and simultaneously for the two components (*fit 9*, *fit 10* and *fit 11* in table 8.12), and it allows to quantify the systematic associated to the choice of the fit configuration.

d) The shape of the continuum background in the fit is assumed to be the same in all bins. However the shapes in the second and third bins are allowed to vary (only in the fit) around the fitted value with Gaussian constraints of mean 1 and width 0.25 and 0.50, respectively. These two variations correspond to the shape differences observed in MC as shown in figure 8.31. This fit configuration is applied to the baseline fit (*fit 1*) and to the two different generations of toy MC experiments described in a). This configuration is labelled *constrained* in table 8.12 (*fit 12*, *fit 13* and *fit 14* in table 8.12).

For completeness, the results obtained for a few configurations which simultaneous fit the B_s^0 and B^0 signals are also shown.

The results obtained with the baseline fit (*fit 1*) are consistent with *fit 12* configuration where the assumption that the continuum background shape is equal in all continuum-BDT bins is relaxed. The difference between the baseline fit and *fit 9* and *fit 10*, where this assumption is dropped for the continuum and SS+SV background respectively, is treated as a systematic uncertainty due to the choice of the fit configuration ($\sim 9\,\%$). Table 8.12 also shows the results obtained when the first bin $0.141 < BDT < 0.252$ is included in the fit. Comparing *fit 17* and *fit 9* we see that the first bin does not improve the signal significance if it is not used to constrain the shape of the continuum background.

Fit	Continuum background shape		SS+SV background shape		B_s^0/B^0	$N(B_s^0)$	$N(B^0)$	B^0		B_s^0	
	Generated	Fitted	Generated	Fitted		gen/bin	gen/bin	Significance	Pull	Significance	Pull
1	= in all bins	= in all bins	= in all bins	= in all bins	SM-like	10	1	not fitted		4.1 ± 1.0	0.12
2	= in all bins	= in all bins	= in all bins	= in all bins	SM-like	10	1	0.9 ± 0.7	-0.06	3.0 ± 1.0	-0.24
3	= in all bins	= in all bins	= in all bins	= in all bins	SM-like	10	1	impose 0.1*$N(B_s^0)$		4.2 ± 1.0	-0.12
4	= in all bins	= in all bins	= in all bins	= in all bins	CMS+LHCb	7	4	1.3 ± 0.9	-0.05	2.0 ± 1.0	-0.13
5	≠ slopes from MC	= in all bins	= in all bins	= in all bins	SM-like	10	1	not fitted		3.9 ± 1.2	0.17
6	≠ slopes from data	= in all bins	= in all bins	= in all bins	SM-like	10	1	not fitted		4.6 ± 1.0	0.36
7	= in all bins	= in all bins	≠ slope from MC	= in all bins	SM-like	10	1	not fitted		4.4 ± 1.0	0.14
8	= in all bins	= in all bins	≠ slope from data	= in all bins	SM-like	10	1	not fitted		3.3 ± 1.1	0.17
9	= in all bins	≠ in each bin	= in all bins	= in all bins	SM-like	10	1	not fitted		3.5 ± 1.3	-0.14
10	= in all bins	= in all bins	= in all bins	≠ in each bin	SM-like	10	1	not fitted		4.0 ± 1.3	0.10
11	= in all bins	≠ in each bin	= in all bins	≠ in each bin	SM-like	10	1	not fitted		3.3 ± 1.0	-0.16
12	= in all bins	constrained	= in all bins	= in all bins	SM-like	10	1	not fitted		4.1 ± 1.1	0.09
13	≠ slopes from MC	constrained	= in all bins	= in all bins	SM-like	10	1	not fitted		3.7 ± 1.1	-0.09
14	≠ slopes from data	constrained	= in all bins	= in all bins	SM-like	10	1	not fitted		4.4 ± 1.0	0.36
4 bins configuration with the bin $0.141 < BDT < 0.252$ included in the fit											
15	= in all bins	= in all bins	= in all bins	= in all bins	SM-like	10	1	not fitted		4.5 ± 1.0	0.18
16	≠ slopes from data	= in all bins	= in all bins	= in all bins	SM-like	10	1	not fitted		4.4 ± 1.0	0.17
17	= in all bins	≠ in each bin	= in all bins	= in all bins	SM-like	10	1	not fitted		3.4 ± 1.0	0.18

Table 8.12. *Summary of the fit results obtained with the different fit configurations.*

8.10.9 Combination of 2011 and 2012 results

The analysis based on 4.9 fb^{-1} of data collected in 2011 resulted in an upper limit of $1.5 \cdot 10^{-8}$ at 95 % CL for the B_s^0 branching ratio (see section 5.10).

The current analysis exploits a larger statistical sample and benefits from important improvements, including better rejection of combinatorial background and of fake muons, a more accurate description of the backgrounds, and the use of the invariant mass fit to extract the signal. The expected sensitivity is now approximately one order of magnitude better than in the previous analysis.

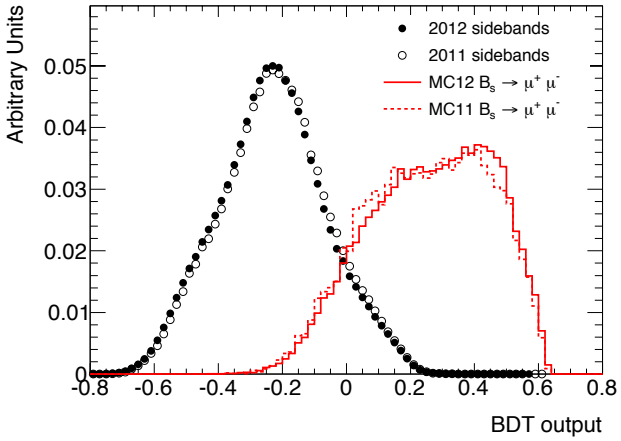


Fig. 8.41. *Continuum-BDT distribution for data sidebands and signal MC events on both 2011 and 2012 samples. The black points (open circles) correspond to 2012 (2011) data sidebands, while the solid (dashed) histogram shows the MC12 (MC11) signal sample.*

In this condition, an average of the two results, based on the combination of the likelihood, is not expected to represent an improvement on the 2012 data alone.

Having completely redesigned the analysis, an important gain on the final significance can be obtained by adding 2011 data to 2012 ones in the final fit. In this direction, 2011 data have been reprocessed with 2012 software and ntuple format, thus including all the improvements (e.g. in the PV association) and the new variables adopted in the fake-BDT as well as in the continuum-BDT. Also the new mass definition, that uses the combined-muons information, is checked on 2011 data. In 2011 events both the "old" (using ID-only information) and the "new" (using the combined-muons information) definitions of the mass are considered. The "new" mass shows a very good agreement with the 2012 distribution. For that reason, for consistency, it was decided to include 2011 data by using the "new" mass definition and all the new selections coherently with 2012 dataset.

The distributions of the discriminating variables used to separate out the combinatorial background have also been checked, and figure 8.41 shows the continuum-BDT distribution for sideband data and signal events. The agreement between 2011 and

2012 is very good in both types of events, even if some small effects are seen in the signal. These have been checked and showed small impact on the continuum-BDT bins efficiency.

In the invariant mass fit 2011 and 2012 data are therefore considered as a one dataset.

8.10.10 Un-blinding

In the previous sections we showed how the invariant mass fit configuration chosen is stable and reliable and how all potential sources of bias and systematic have been explored and estimated.

At the time of the thesis submission the $B_s^0 \to \mu^+\mu^-$ analysis was still in progress and refined tests of the signal selection and background rejection were necessary before the publication of the final results.

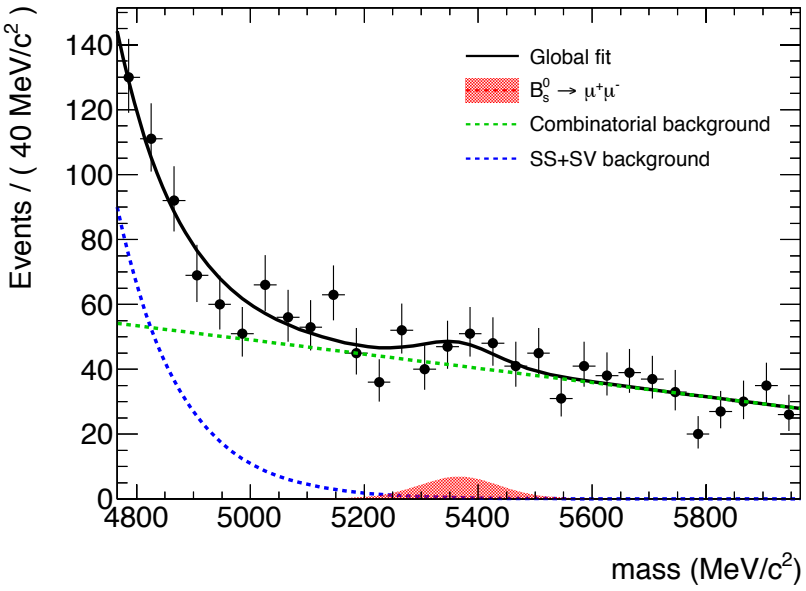


Fig. 8.42. *Un-binned maximum likelihood fit performed on 2012 un-blinded data in the BDT bin* $[0.252, 1.0]$. *The combinatorial shape (green dashed line) is fixed to the value obtained from the simultaneous fit of the three BDT bins reported in the section 8.10.4, while the exponential for the low-mass background (blue dashed line) is floating.*

Figure 8.42 reports the invariant mass fit performed on Run 1 un-blinded data using the continuum-BDT bin $[0.252, 1.0]$ sum of the three optimised for the 2012 analysis and discussed in section 8.10.1. The yield obtained from this non-optimised version of the fit is 32 ± 21, in agreement, although the big statistical error, with the number of events expected from ATLAS Run 1 data.

8.11 Conclusions and Future Prospects

In this final chapter of my thesis I showed how all the studies I performed during my Ph.D. proved to be fundamental for the $B_s^0 \to \mu^+\mu^-$ measurement performed on the full Run 1 data. The increased statistic along with various improvements in the analysis selection and the introduction of an invariant mass fit for the extraction of the signal yield, allowed to increase the expected sensitivity with respect to the previous public analysis of about a factor ten. Despite the ATLAS detector weaknesses in the trigger and in the mass resolution, the sensitivity expected on 25 fb^{-1} of integrated luminosity is approaching the level of LHCb and CMS.

The analysis still have to be finalised after the problem observed on the un-blinded data, but several test are being performed and the results are promising.

Of course, the ATLAS $B_s^0 \to \mu^+\mu^-$ search will continue also during the second period of data taking benefiting from the increased luminosity and the increased energy provided by LHC.

In order to have a rough estimation of the performance expected with Run 2 data (and beyond), the expected sensitivity for the $B_s^0 \to \mu^+\mu^-$ channel has been evaluated using the baseline fit described in this chapter. Figure 8.43 shows the ATLAS expected significance as function of the integrated luminosity obtained assuming that the background scales as the signal with the luminosity. Considering the current ATLAS detector and reconstruction performance, the expected sensitivity for ~ 100 fb^{-1} of data is about 7 σ.

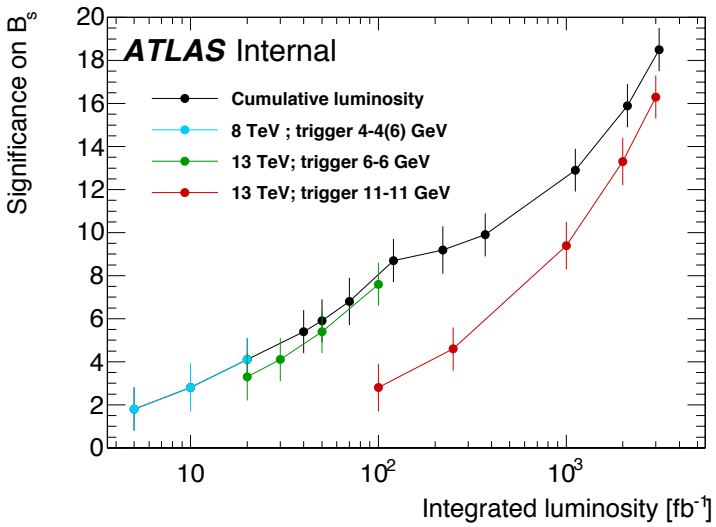


Fig. 8.43. *Expected $B_s^0 \to \mu^+\mu^-$ significance as a function of the integrated luminosity.*

During the first long shutdown the ATLAS inner tracker system has been updated with the introduction of a new layer of pixel closer to the beam line that will improve the measurement of the impact parameter and the mass resolution. This is extremely

important for the $B_s^0 \to \mu^+\mu^-$ analysis since it will allow to increase the separation between B_s^0 and B^0 mass distributions improving therefore the sensitivity on both channels.

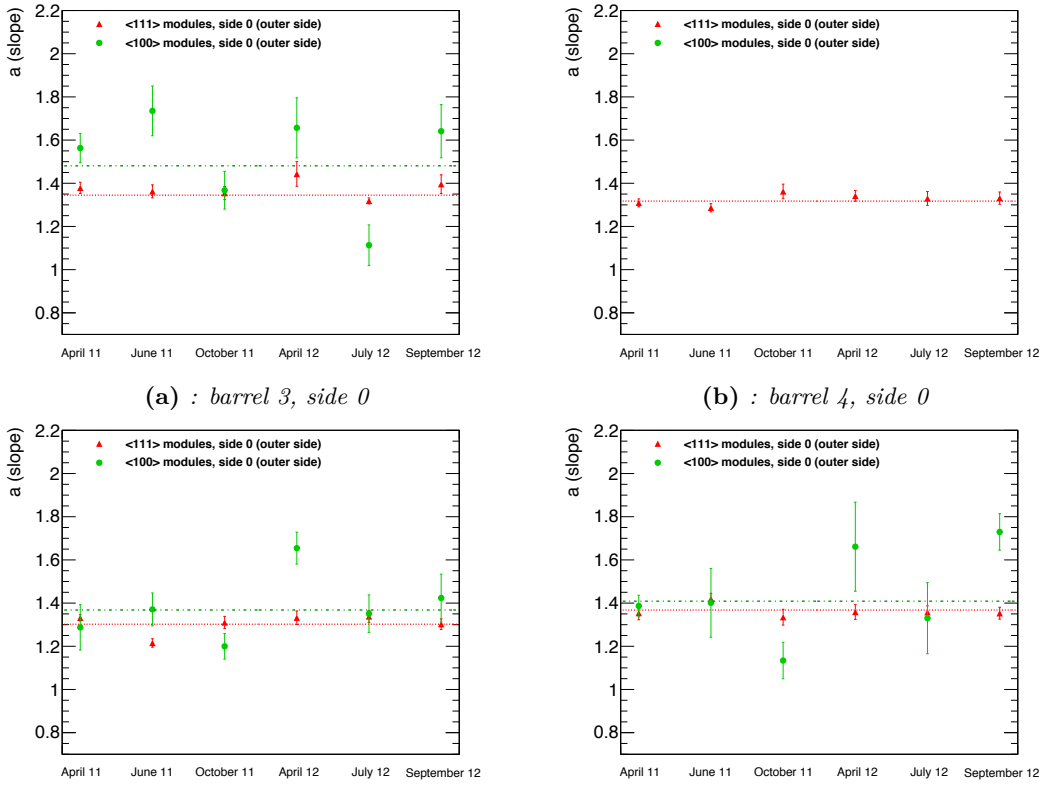
In addition, the ATLAS trigger system has been upgraded, introducing further selection both at L1 and HLT that will allow to use low muon thresholds (6 GeV) even in the high pile-up scenario expected in Run 2.

Appendix A

Additional Lorentz Angle Fit Tests

A.1 Behaviour of the Fit Parameters in 2011 and 2012

Figures A.1, A.2, A.3, A.4, A.5 and A.6 show respectively the behaviour of the slope parameter a, the minimum cluster width b and the smearing term σ during 2011 and 2012 data taking period for side 0 and side 1 of the modules. The dotted red and dash-dotted green horizontal lines in the plots show the mean value of the parameters in 2011 and 2012 for <111> and <100> modules respectively.



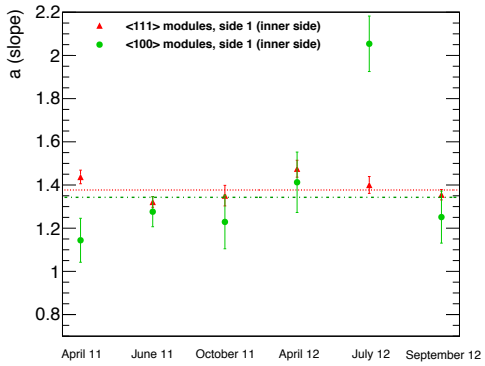
(a) *: barrel 3, side 0*

(b) *: barrel 4, side 0*

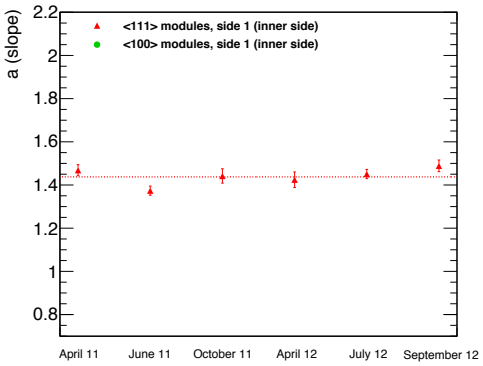
(c) *: barrel 5, side 0*

(d) *: barrel 6, side 0*

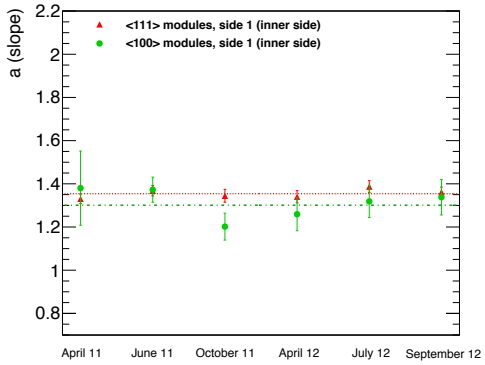
Fig. A.1. *Slope parameter a in 2011 and 2012 for side 0 of the modules.*



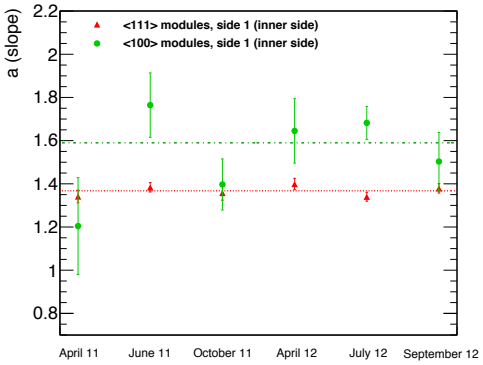
(a) *: barrel 3, side 1*



(b) *: barrel 4, side 1*

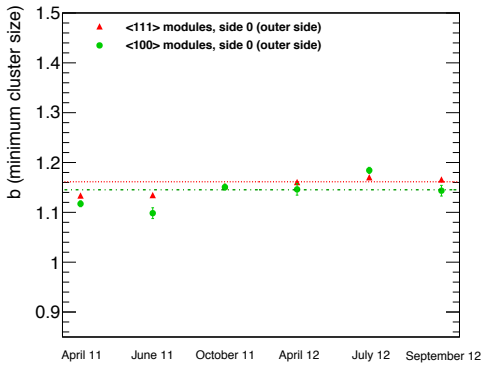


(c) *: barrel 5, side 1*

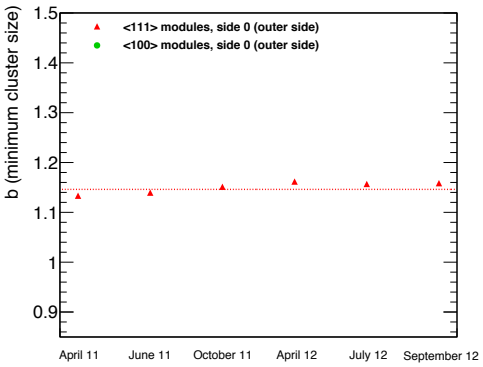


(d) *: barrel 6, side 1*

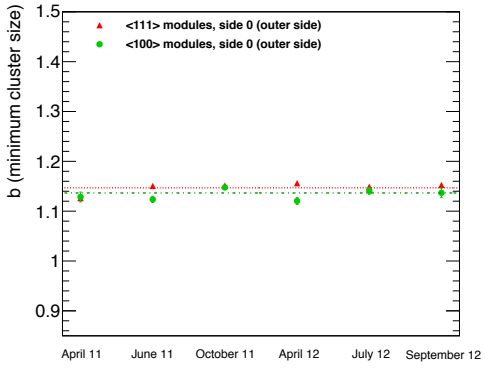
Fig. A.2. *Slope parameter a in 2011 and 2012 for side 1 of the modules.*



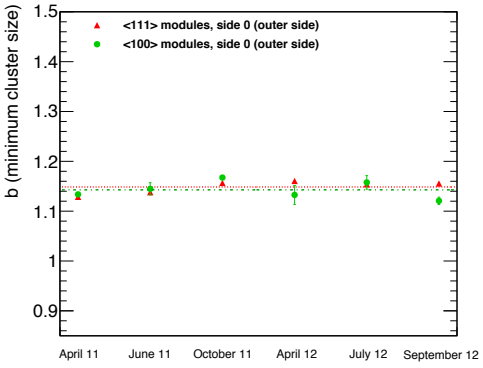
(a) *: barrel 3, side 0*



(b) *: barrel 4, side 0*

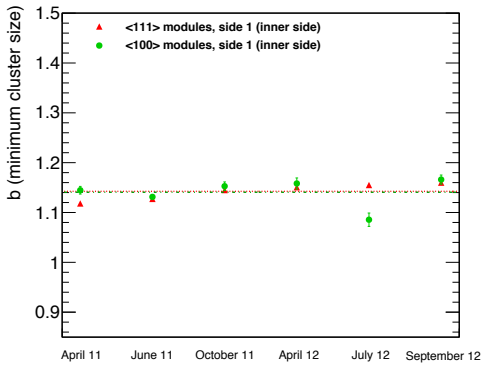


(c) *: barrel 5, side 0*

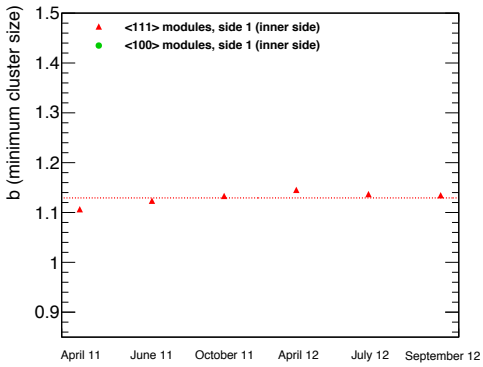


(d) *: barrel 6, side 0*

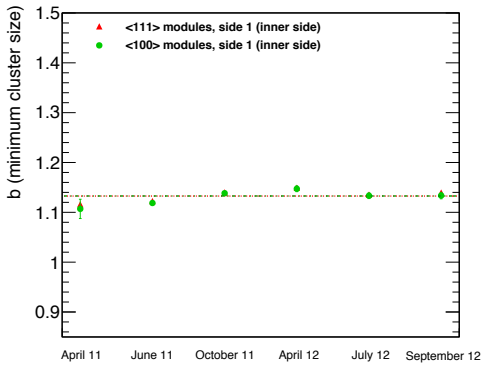
Fig. A.3. *Cluster width b in 2011 and 2012 for side 0 of the modules.*



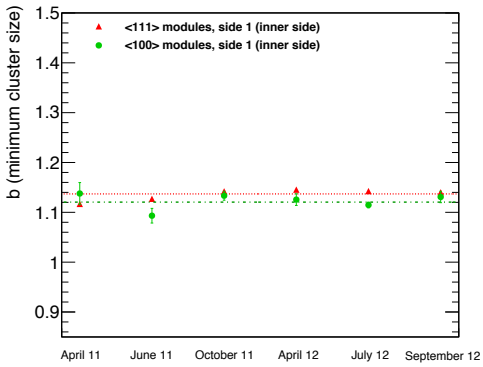
(a) *: barrel 3, side 1*



(b) *: barrel 4, side 1*

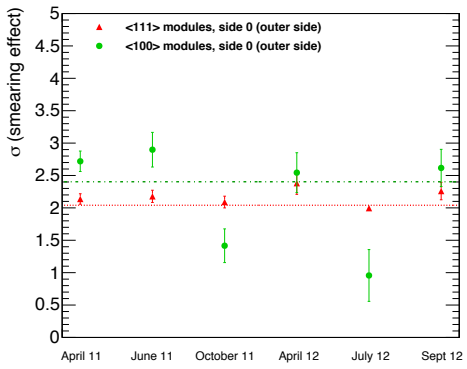


(c) *: barrel 5, side 1*

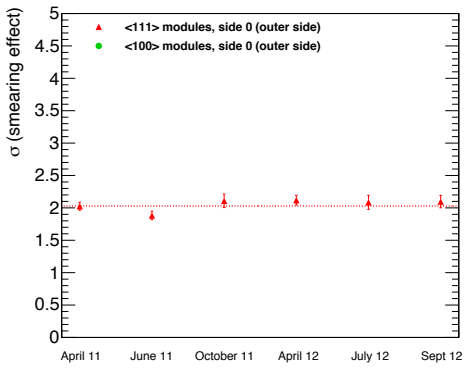


(d) *: barrel 6, side 1*

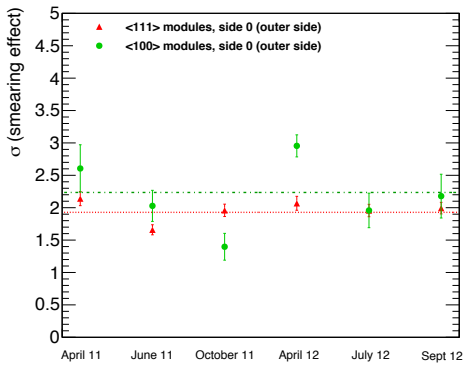
Fig. A.4. *Cluster width b in 2011 and 2012 for side 1 of the modules.*



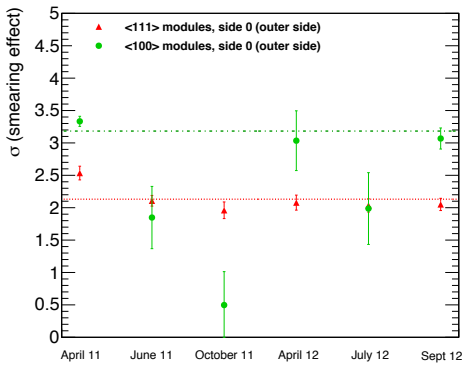
(a) *: barrel 3, side 0*



(b) *: barrel 4, side 0*

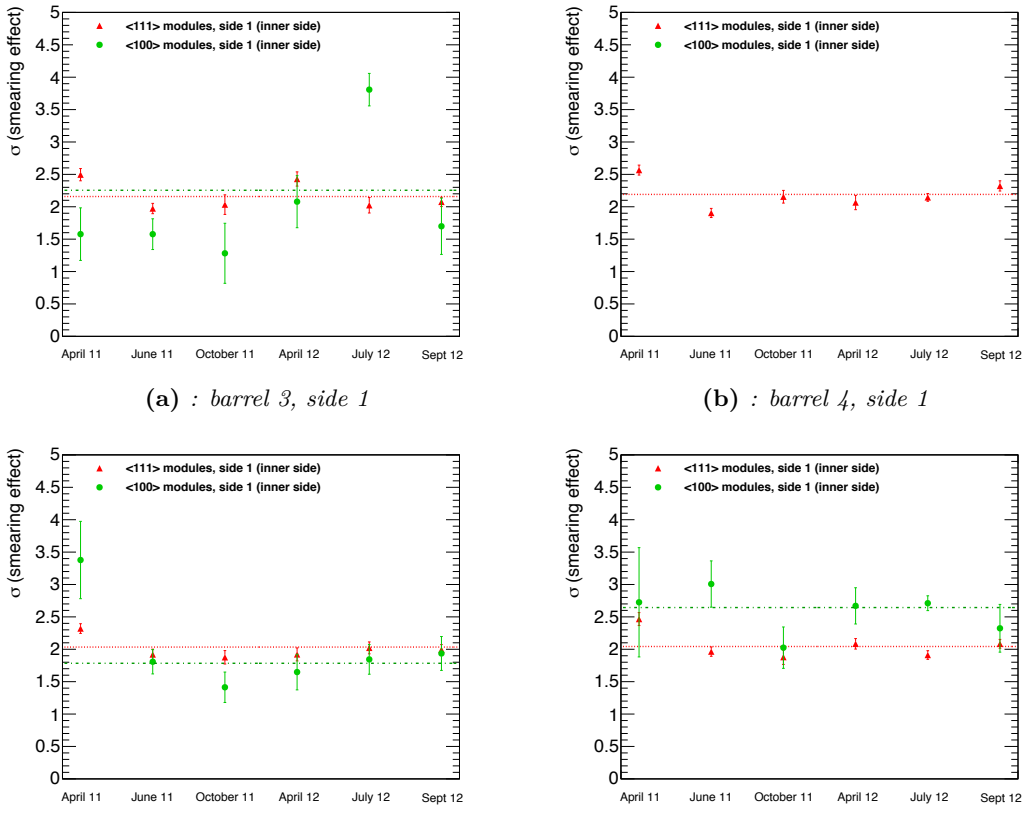


(c) *: barrel 5, side 0.25*



(d) *: barrel 6, side 0*

Fig. A.5. *Smearing factor σ in 2011 and 2012 for side 0 of the modules.*



(a) *: barrel 3, side 1*

(b) *: barrel 4, side 1*

(c) *: barrel 5, side 1*

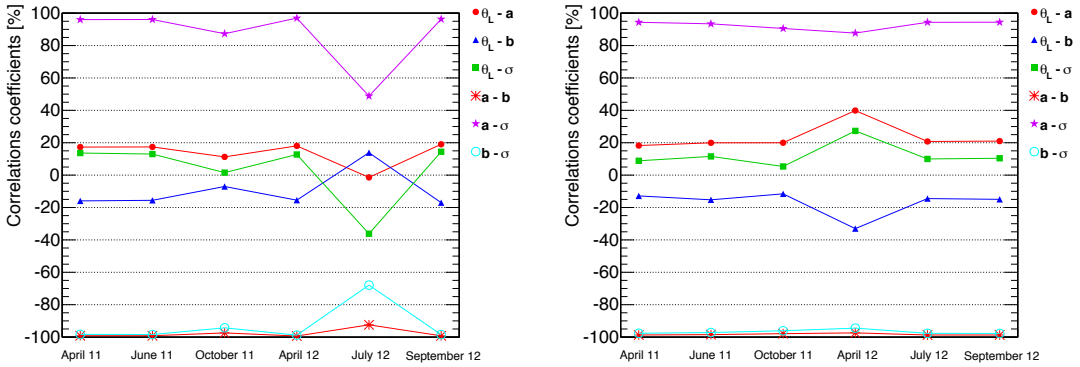
(d) *: barrel 6, side 1*

Fig. A.6. *Smearing factor σ in 2011 and 2012 for side 1 of the modules.*

The fit parameters for <111> modules are stable along the data taking runs. Statistical fluctuations in the slope parameter a and the smearing term σ are present for <100> modules, due to the lower statistic.

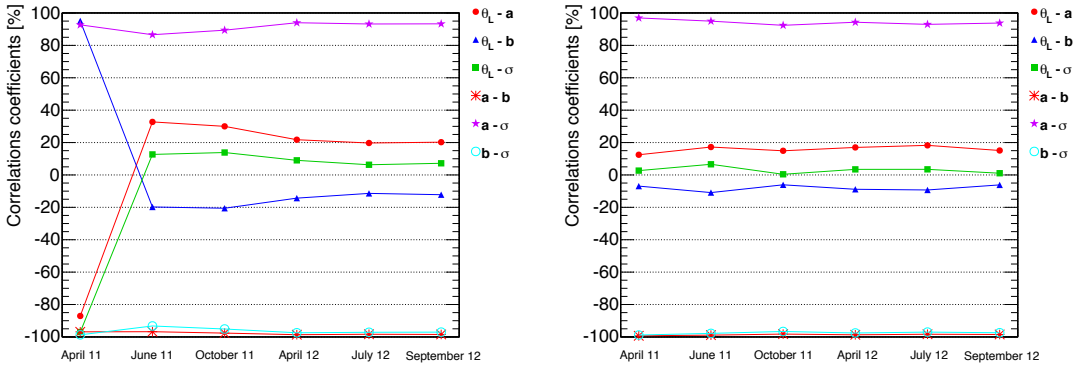
A.2 Fit Parameters Correlation in 2011 and 2012

Figures A.7, A.8, A.9 and A.10 show the correlation of the fit parameters during 2011 and 2012 data taking period for <111> and <100> modules. The Lorentz angle is not strongly correlated with the other three parameters, while a, b and σ are highly correlated among each other.



(a) *: barrel 3, side 0*

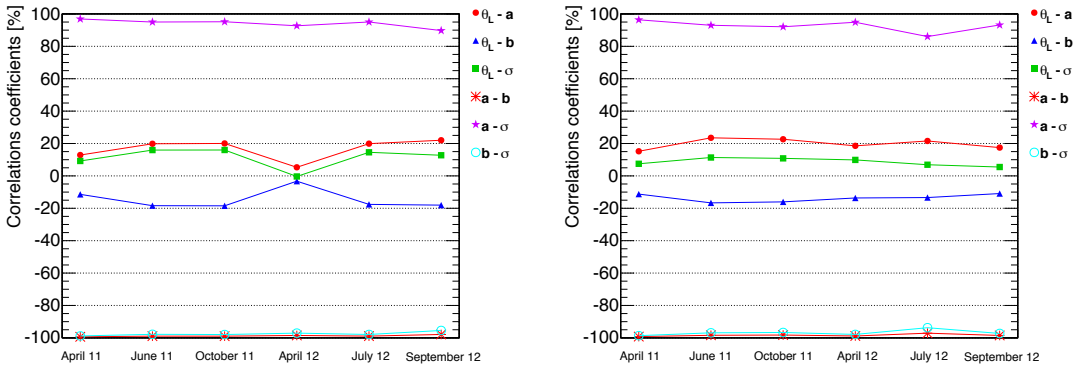
(b) *: barrel 4, side 0*



(c) *: barrel 5, side 0*

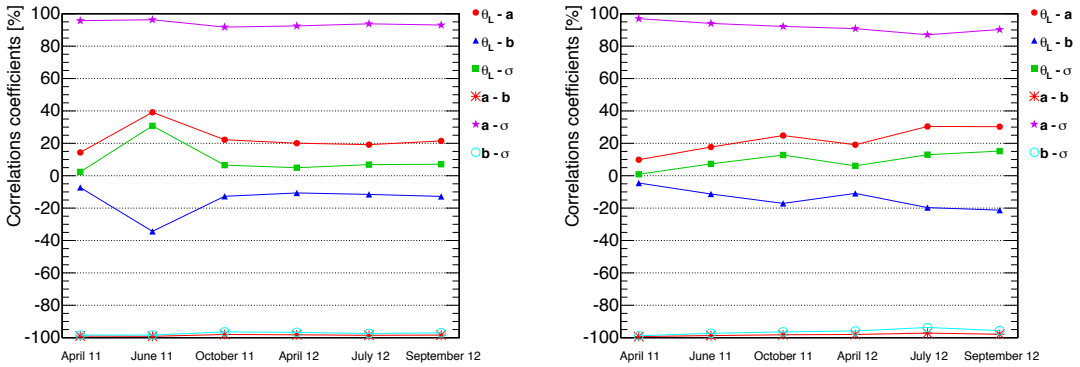
(d) *: barrel 6, side 0*

Fig. A.7. *Correlation of the fit parameters for <111> modules (side 0) during 2011 and 2012.*



(a) *: barrel 3, side 1*

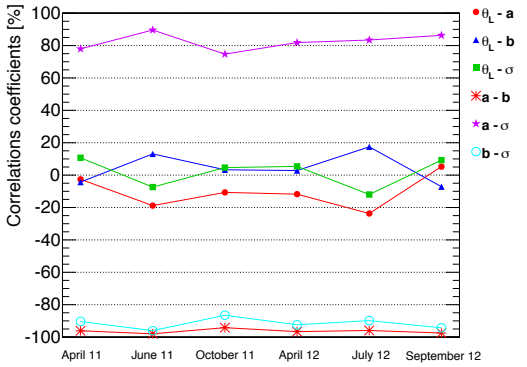
(b) *: barrel 4, side 1*



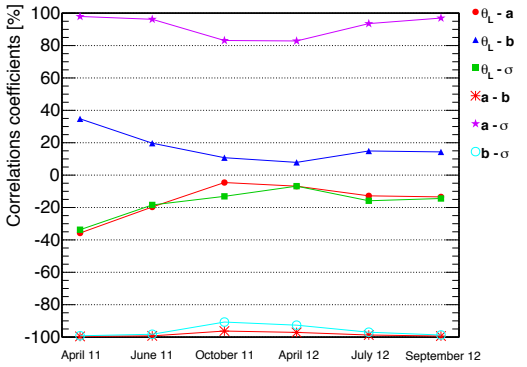
(c) *: barrel 5, side 1*

(d) *: barrel 6, side 1*

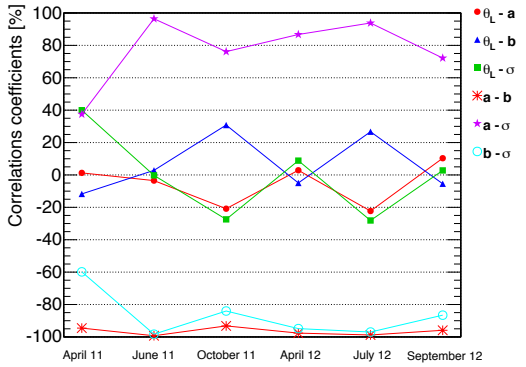
Fig. A.8. *Correlation of the fit parameters for <111> modules (side 1) during 2011 and 2012.*



(a) *: barrel 3, side 0*

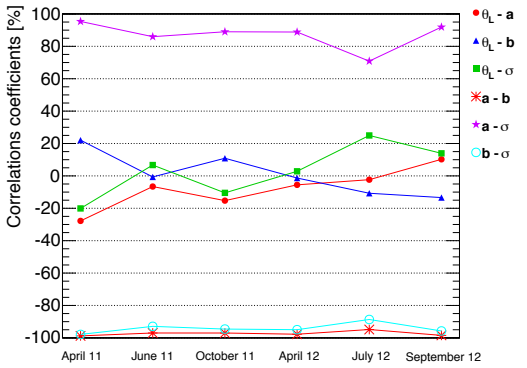


(b) *: barrel 5, side 0*

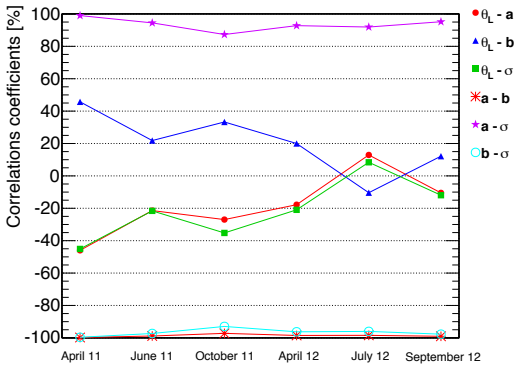


(c) *: barrel 6, side 0*

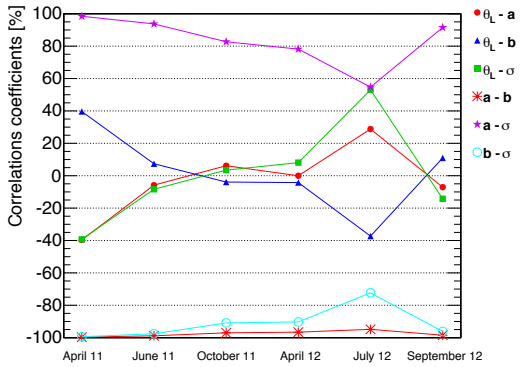
Fig. A.9. *Correlation of the fit parameters for <100> modules (side 0) during 2011 and 2012.*



(a) *: barrel 3, side 1*



(b) *: barrel 5, side 1*



(c) *: barrel 6, side 1*

Fig. A.10. *Correlation of the fit parameters for <100> modules (side 1) during 2011 and 2012.*

A.3 Binning Effects in Profile Histograms

A possible bias on the fitted value of the Lorentz angle, introduced by the choice of the profile histogram binning, has been checked varying the default bin width. The results showed in chapter 3 have been obtained considering a binning of 0.5°. Three other values (0.25°, 1.0° and 2.0°) has been tested for both <111> and <100> modules. The results of the fits (for side 1) are shown in figures A.11 and A.12, while the respective Lorentz angle values are summarised tables in A.1 and A.2.

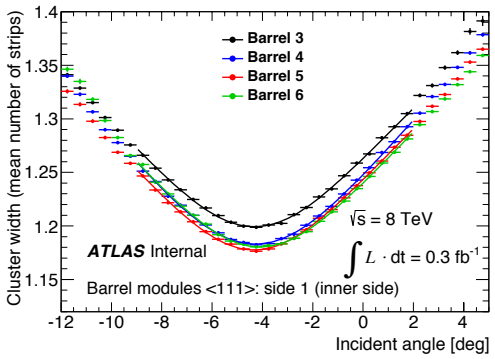
Barrel	0.5°	0.25°	1.0°	2.0°
3	- 4.33 ± 0.01	- 4.33 ± 0.01	- 4.33 ± 0.01	- 4.34 ± 0.01
4	- 4.29 ± 0.01	- 4.30 ± 0.01	- 4.29 ± 0.01	- 4.31 ± 0.01
5	- 4.34 ± 0.01	- 4.34 ± 0.01	- 4.35 ± 0.01	- 4.38 ± 0.01
6	- 4.11 ± 0.01	- 4.11 ± 0.01	- 4.12 ± 0.01	- 4.14 ± 0.01

Table A.1. *Lorentz angle values obtained considering different TProfile bin widths for <111> modules (side 1). The width 0.5° corresponds to the default binning.*

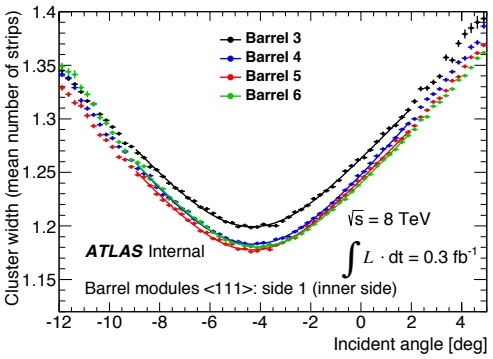
Barrel	0.5°	0.25°	1.0°	2.0°
3	- 3.32 ± 0.05	- 3.30 ± 0.05	- 3.32 ± 0.07	- 3.31 ± 0.04
4	-	-	-	-
5	- 3.43 ± 0.03	- 3.43 ± 0.03	- 3.44 ± 0.03	- 3.45 ± 0.03
6	- 3.28 ± 0.05	- 3.29 ± 0.05	- 3.31 ± 0.06	- 3.31 ± 0.08

Table A.2. *Lorentz angle values obtained considering different TProfile bin widths for <100> modules (side 1). The width 0.5° corresponds to the default binning.*

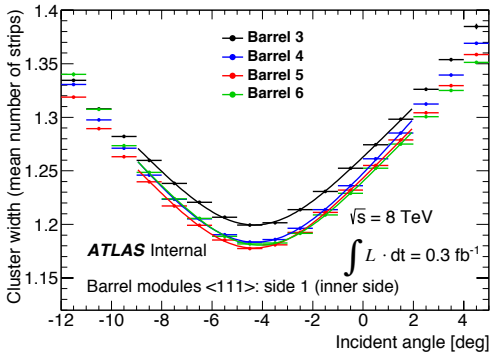
The results show how the minimum of the cluster width distribution is independent (within the statistical error) of the width of the bin used to make the TProfile plots. The same behaviour has been found for side 0 of the modules.



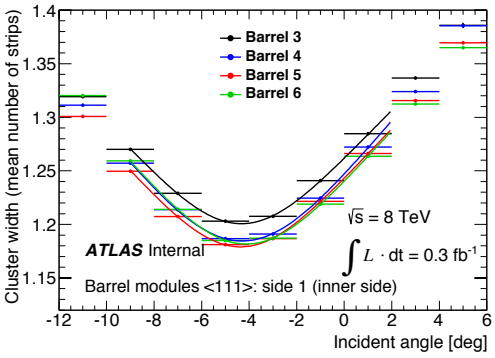
(a) : 0.5° *(default binning)*



(b) : 0.25°

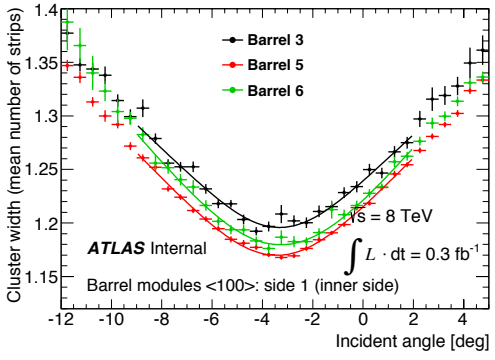


(c) : 1.0°

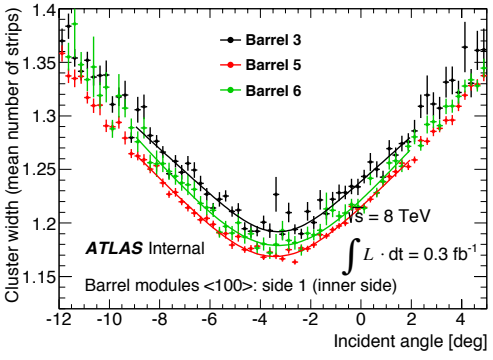


(d) : 2.0°

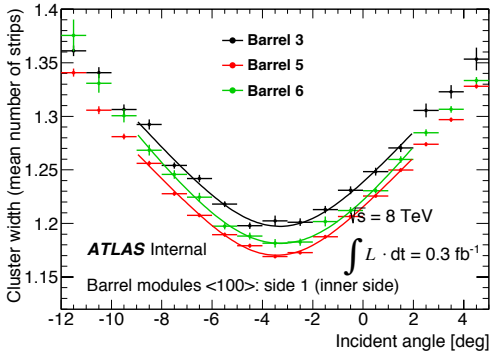
Fig. A.11. *Lorentz angle fit for different TProfile bin widths (<111> modules, side 1).*



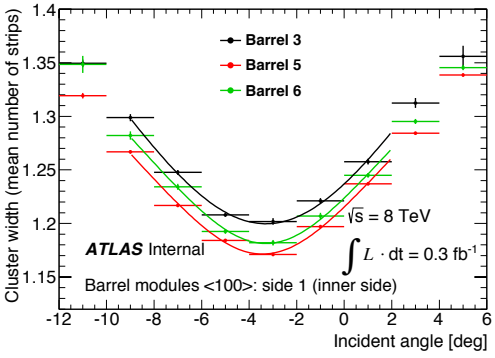
(a) : 0.5° *(default binning)*



(b) : 0.25°



(c) : 1.0°



(d) : 2.0°

Fig. A.12. *Lorentz angle fit for different TProfile bin widths (<100> modules, side 1).*

A.4 Effects of the δ-rays on the Fit Performance

Charged particles can deposit energy in silicon leading to the production of secondary electrons, called δ-rays, that can travel several hundred microns and produce secondary ionisation. These δ-rays can create hits in the neighbouring strips increasing the cluster width and, therefore, affecting the mean value of the Lorentz angle. For that reason, the fit has been performed applying three different cuts, as shown in figure A.13. The results of the fit (for side 1) are shown in figures A.14 and A.15, while the respective Lorentz angle values are summarised in tables A.3 and A.4.

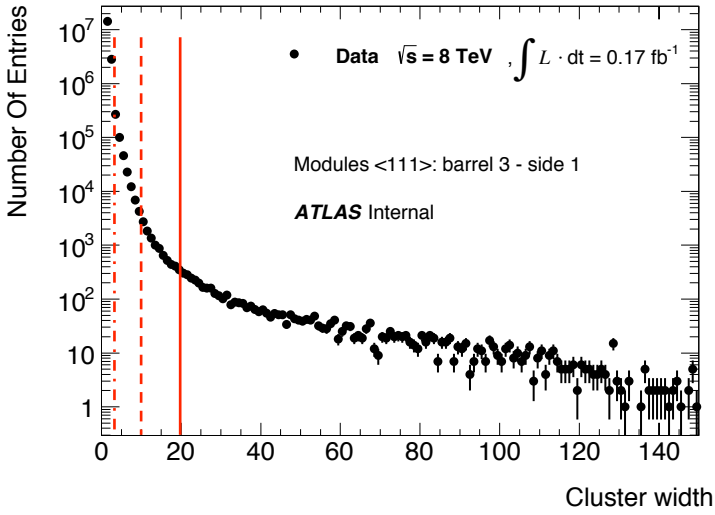


Fig. A.13. *Cluster width distribution for <111> modules (barrel 3, side 1). The solid, dashed and dash-dotted red vertical lines show the cuts applied on the cluster width at 20, 10 and 3 respectively.*

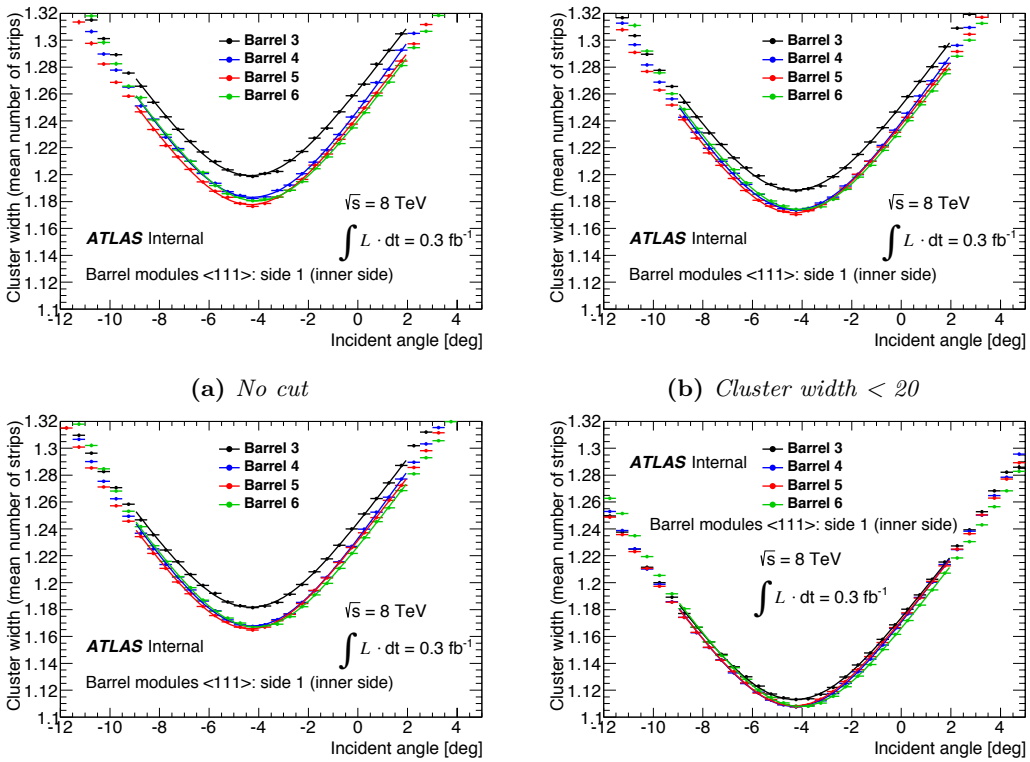
Barrel	No cut	Cluster width < 20	Cluster width < 10	Cluster width < 3
3	- 4.33 ± 0.01	- 4.33 ± 0.01	- 4.34 ± 0.01	- 4.302 ± 0.004
4	- 4.29 ± 0.01	- 4.29 ± 0.01	- 4.29 ± 0.01	- 4.302 ± 0.004
5	- 4.34 ± 0.01	- 4.34 ± 0.01	- 4.35 ± 0.01	- 4.331 ± 0.004
6	- 4.11 ± 0.01	- 4.13 ± 0.01	- 4.13 ± 0.01	- 4.105 ± 0.005

Table A.3. *Lorentz angle values obtained for different cuts on the cluster width (<111> modules, side 1). The default fit is performed without any cut.*

The cuts on the tail of the cluster width distribution change, as expected, the mean size of the cluster in each ϕ bin of the profile plot, without affecting the local minimum of the distribution. This further proves the stability of the Lorentz angle results. In addition, by removing the tail, the statistical error of the fit decreases. The same behaviour has been found for side 0.

Barrel	No cut	Cluster width < 20	Cluster width < 10	Cluster width < 3
3	- 3.32 ± 0.05	- 3.36 ± 0.06	- 3.36 ± 0.03	- 3.29 ± 0.02
4	-	-	-	-
5	- 3.43 ± 0.03	- 3.43 ± 0.03	- 3.44 ± 0.02	- 3.42 ± 0.01
6	- 3.28 ± 0.05	- 3.30 ± 0.06	- 3.30 ± 0.04	- 3.27 ± 0.03

Table A.4. *Lorentz angle values obtained for different cuts on the cluster width (<100> modules, side 1). The default fit is performed without any cut.*



(a) *No cut*

(b) *Cluster width < 20*

(c) *Cluster width < 10*

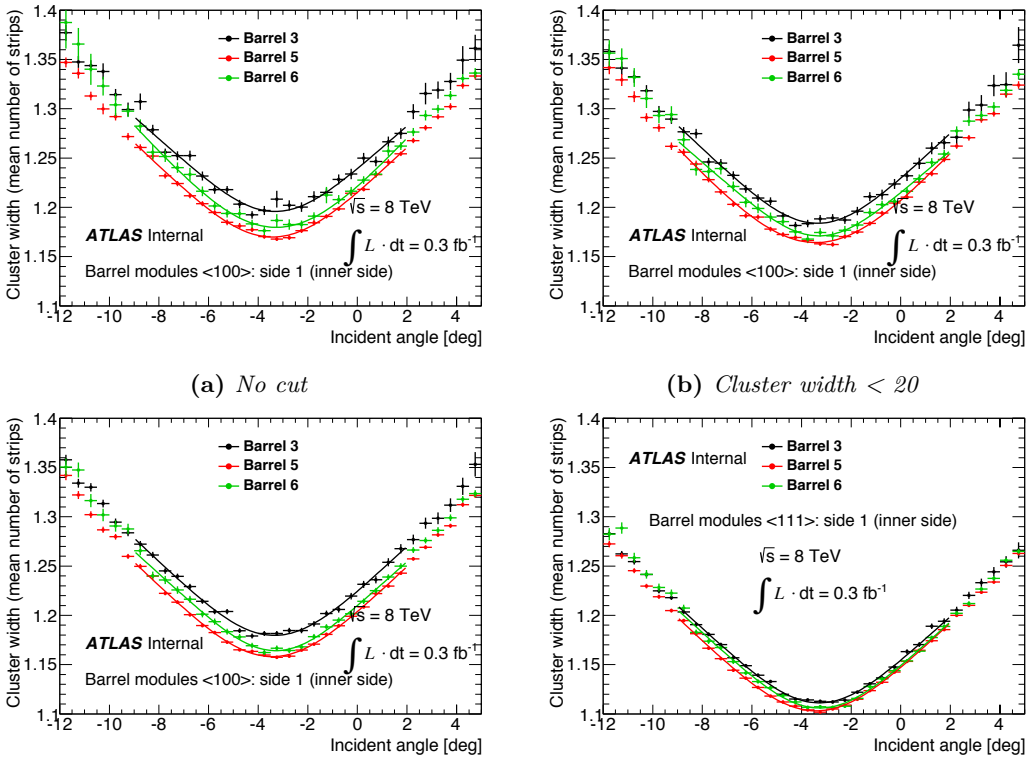
(d) *Cluster width < 3*

Fig. A.14. *Lorentz angle fit for different cuts on the cluster width (<111> modules, side 1).*

A.5 Fit Using a Parabola

In order to quantify the systematic associated to the choice of the fit function, the Lorentz angle has been extracted modelling the cluster width distribution using a parabola defined as

$$f(\phi) = a + b \cdot \phi + c \cdot \phi^2 \quad . \tag{A.1}$$



(a) *No cut*

(b) *Cluster width < 20*

(c) *Cluster width < 10*

(d) *Cluster width < 3*

Fig. A.15. *Lorentz angle fit for different cuts on the cluster width (<100> modules, side 1).*

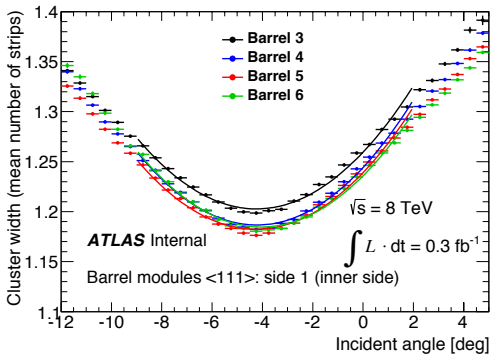
Consequently, the minimum of the distribution, corresponding to the Lorenz angle, is

$$\frac{df(\phi)}{d\phi} = 0 \;\rightarrow\; \phi_{\min} \equiv \phi_L = -\frac{b}{2c} \quad , \tag{A.2}$$

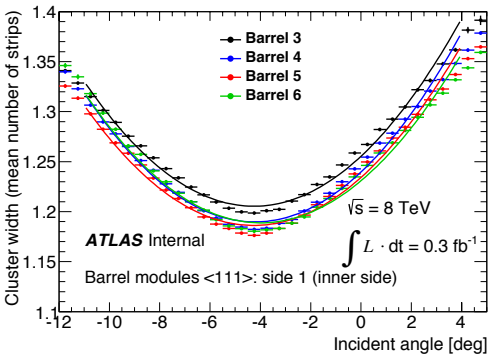
with an error of

$$\Delta\phi_L = \sqrt{\left(-\frac{1}{2c}\cdot\Delta b\right)^2 + \left(\frac{b}{2c^2}\cdot\Delta c\right)^2 + 2\cdot\left(-\frac{b}{4c^3}\right)^2\cdot \mathrm{COV}[b,c]} \quad . \tag{A.3}$$

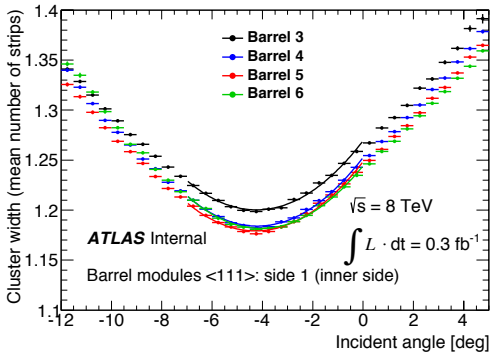
Figures A.16 and A.17 show the results of the fit performed on <111> and <100> modules (side 1) with different fit ranges. Unfortunately, the parabola is not able to properly model the cluster width distribution and therefore can not be used in the estimation of this kind of systematic.



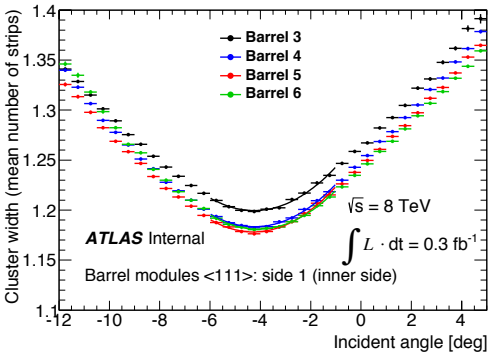
(a) $-9 < \phi < +2$ *(default range)*



(b) $-11 < \phi < +4$

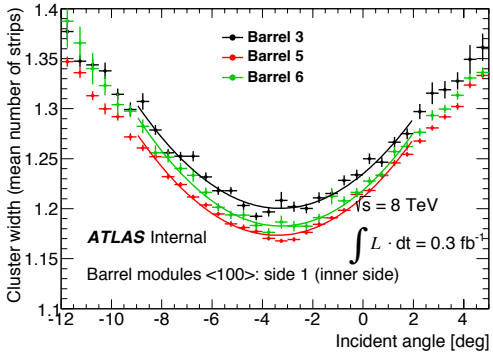


(c) $-7 < \phi < 0$

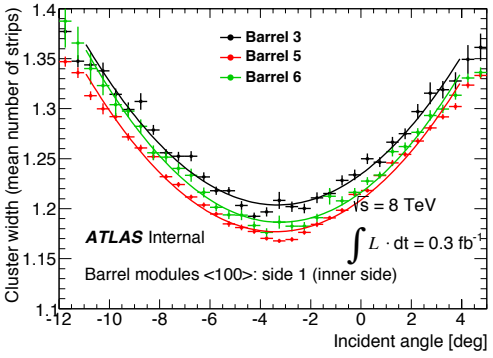


(d) $-6 < \phi < -1$

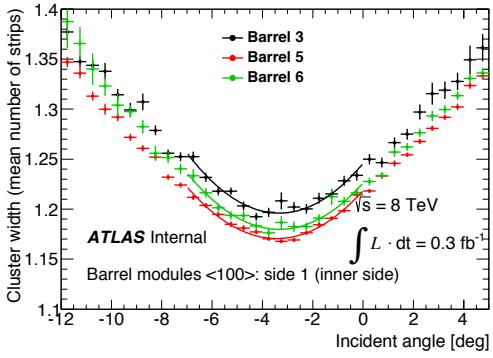
Fig. A.16. *Lorentz angle fit performed using a parabola (<111> modules, side 1).*



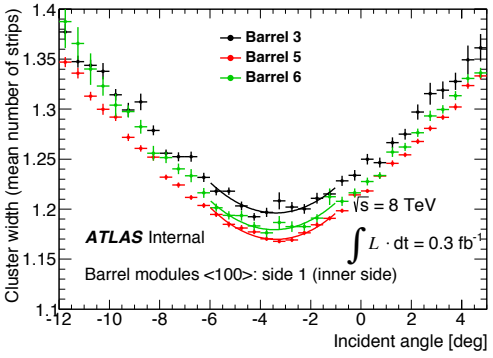
(a) $-9 < \phi < +2$ *(default range)*



(b) $-11 < \phi < +4$



(c) $-7 < \phi < 0$



(d) $-6 < \phi < -1$

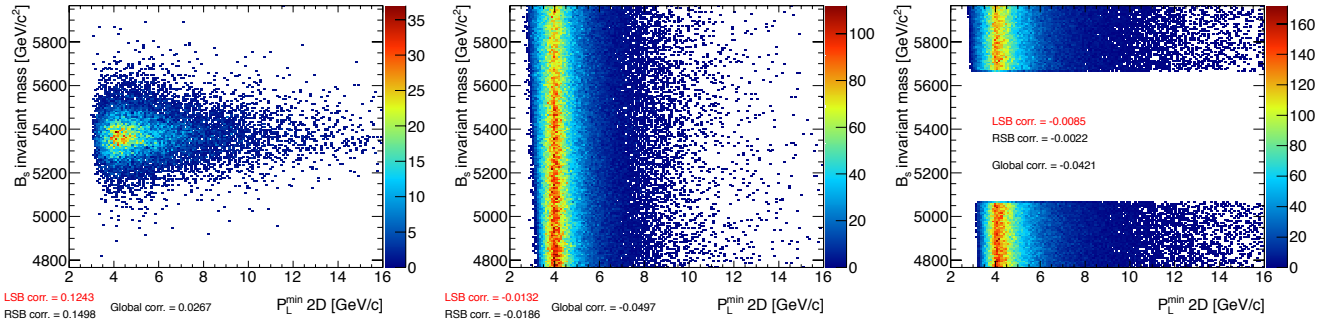
Fig. A.17. *Lorentz angle fit performed using a parabola (<100> modules, side 1).*

Appendix B

Discriminating Variables

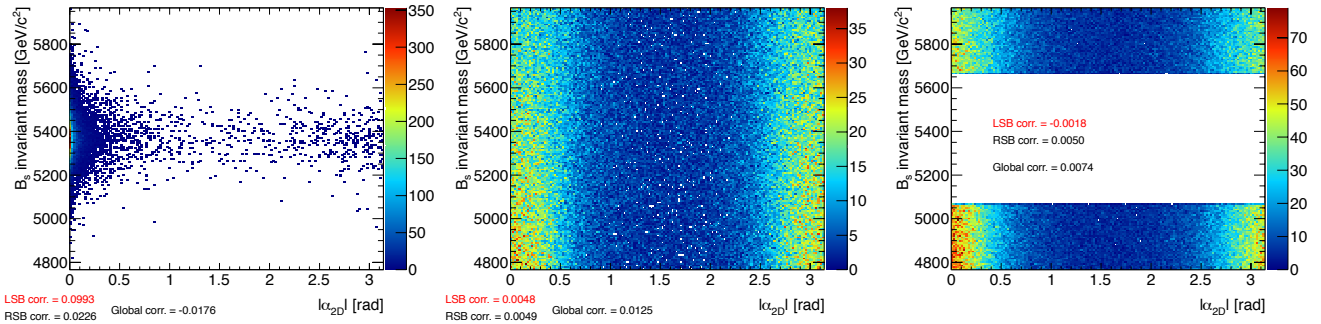
B.1 Mass Correlation of the Discriminating Variables

The following figures report the scatter plots (B_s mass - discriminating variables) for all the discriminating variable used in the 2011 analysis and detailed in section 5.5.1.

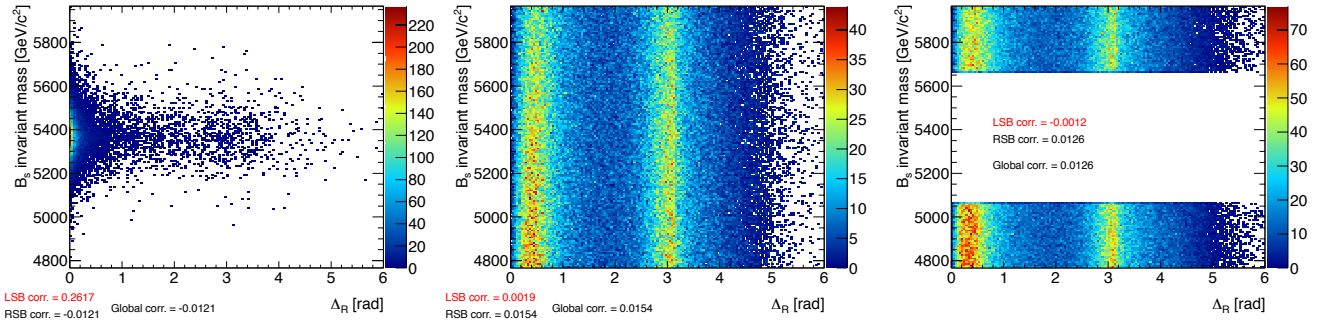


(a) p_L^{min}

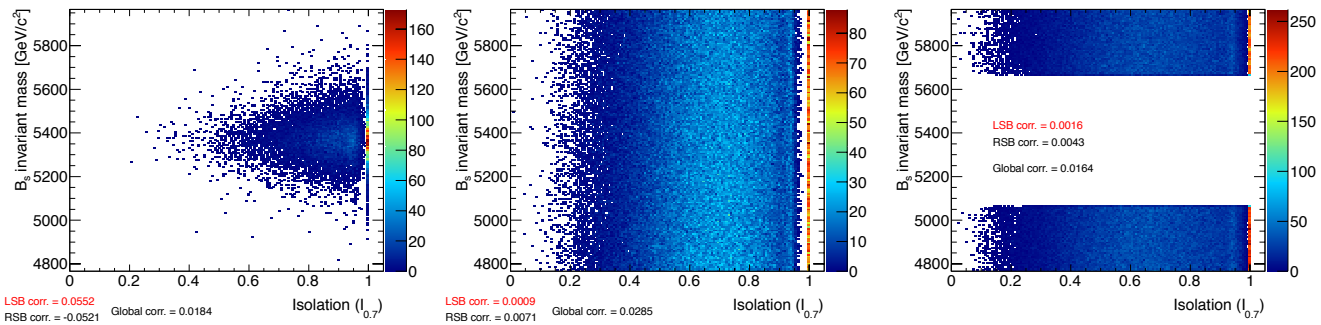
Fig. B.1. *Scatter plot (B_s mass - discriminating variables) for $B_s^0 \to \mu^+\mu^-$ signal (left), $b\bar{b} \to \mu^+\mu^- X$ background (middle) and 2011 un-blinded data (right) for the discriminating variable p_L^{min}. The global correlation, as well as the correlation on the right (RSB) and left (LSB) sidebands is reported in the plots.*



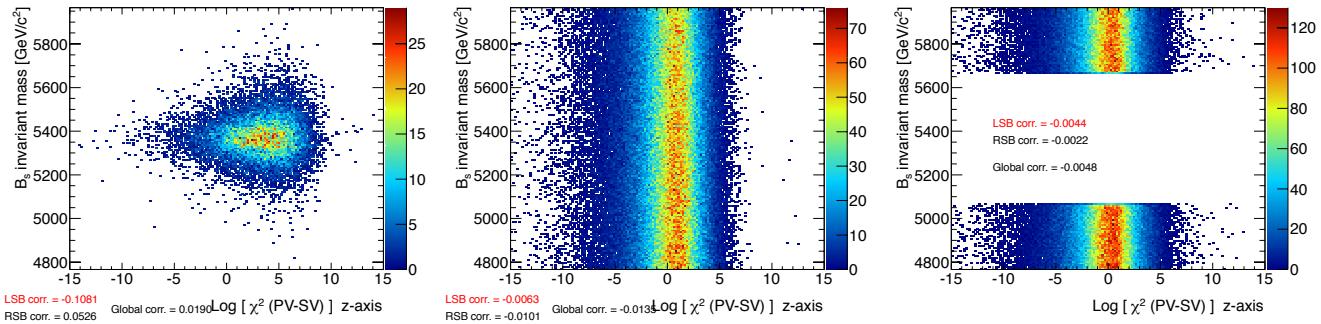
(a) *Pointing angle* $|\alpha_{2D}|$



(b) ΔR

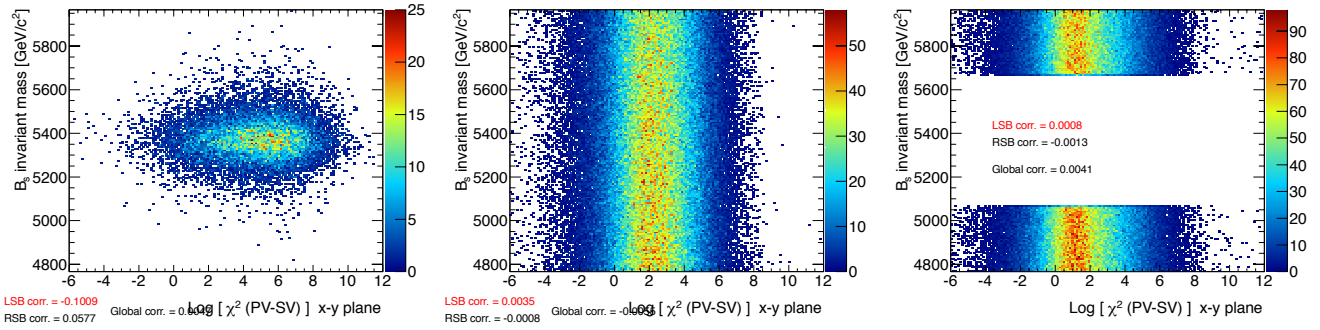


(c) *Isolation* $(I_{0.7})$

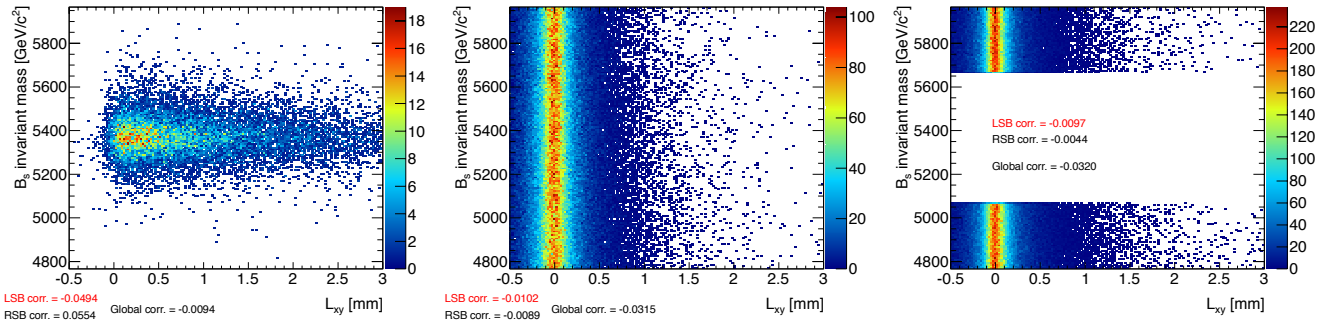


(d) $\chi^2(PV - SV)$ *along z-axis*

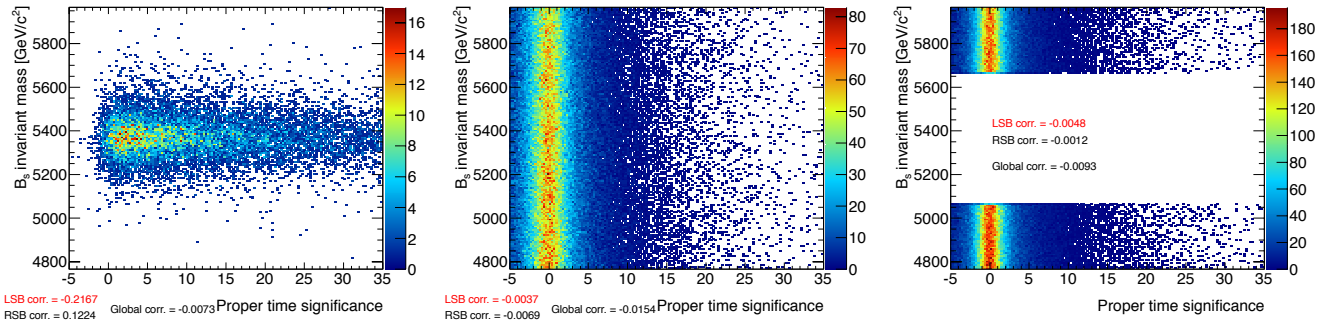
Fig. B.2. *Scatter plot (B_s mass - discriminating variables) for $B_s^0 \to \mu^+\mu^-$ signal (left), $b\bar{b} \to \mu^+\mu^- X$ background (middle) and 2011 un-blinded data (right) for the discriminating variables α_{2D}, ΔR, Isolation and $\chi^2(PV - SV)$ along z-axis. The global correlation, as well as the correlation on the right (RSB) and left (LSB) sidebands is reported in the plots.*



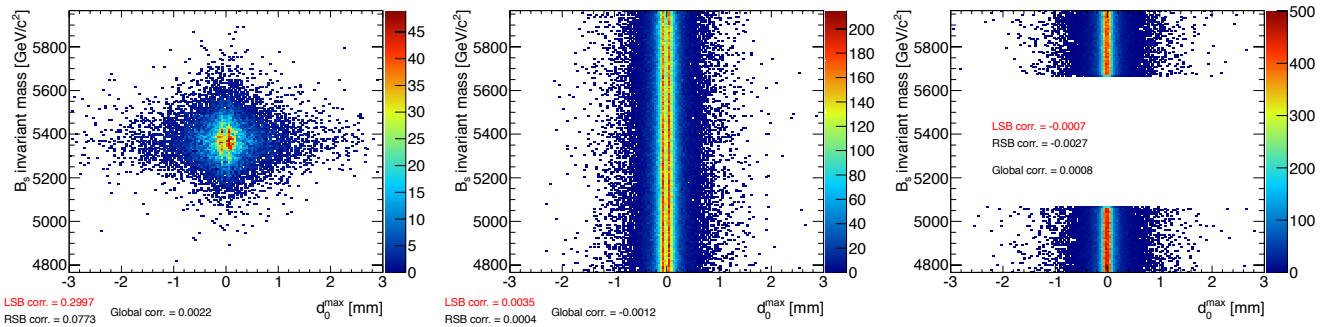
(a) $\chi^2(PV-SV)$ *in the x-y plane*



(b) L_{xy}

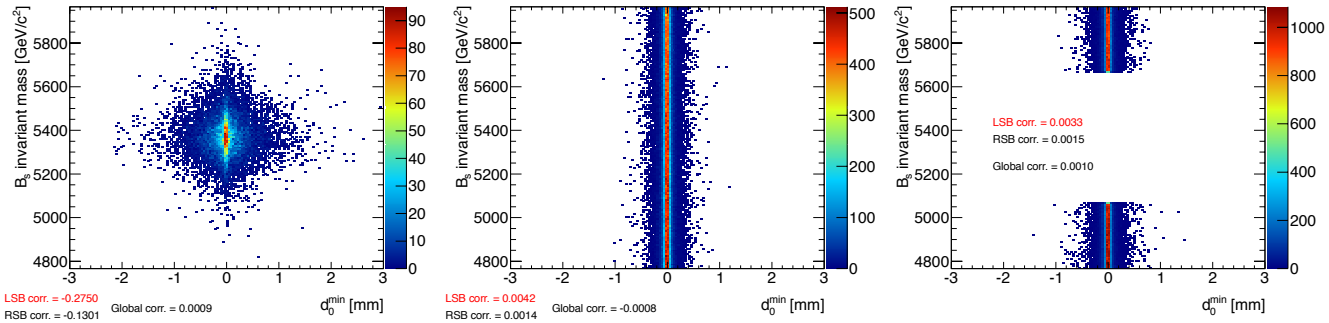


(c) *Proper time significance*

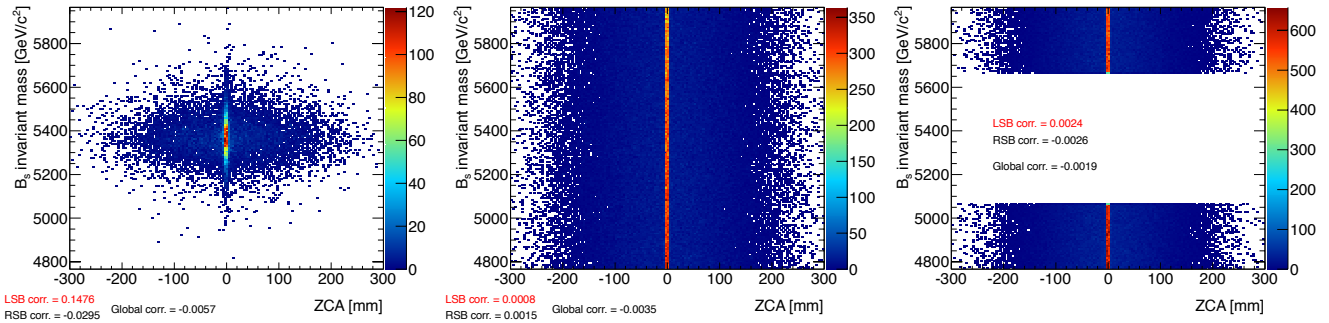


(d) d_0^{max}

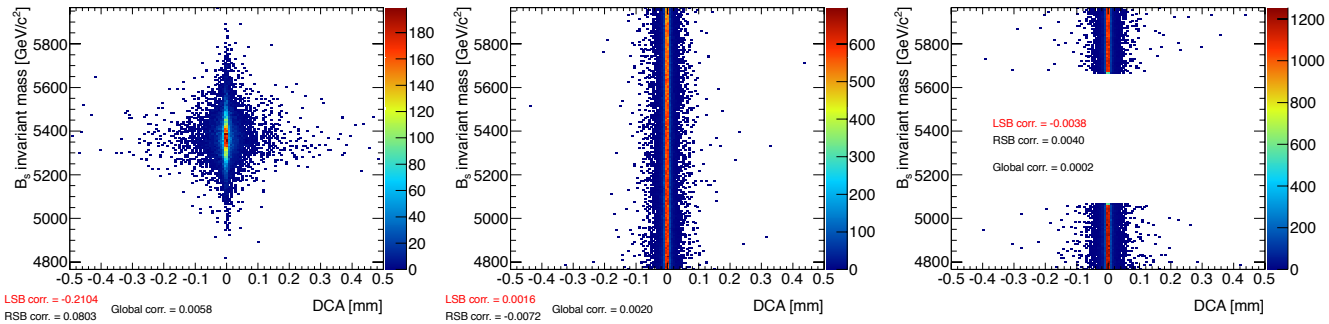
Fig. B.3. *Scatter plot (B_s mass - discriminating variables) for $B_s^0 \to \mu^+\mu^-$ signal (left), $b\bar{b} \to \mu^+\mu^- X$ background (middle) and 2011 un-blinded data (right) for the discriminating variables $\chi^2(PV-SV)$ in the x-y plane, L_{xy}, proper time significance and d_0^{max}. The global correlation, as well as the correlation on the right (RSB) and left (LSB) sidebands is reported in the plots.*



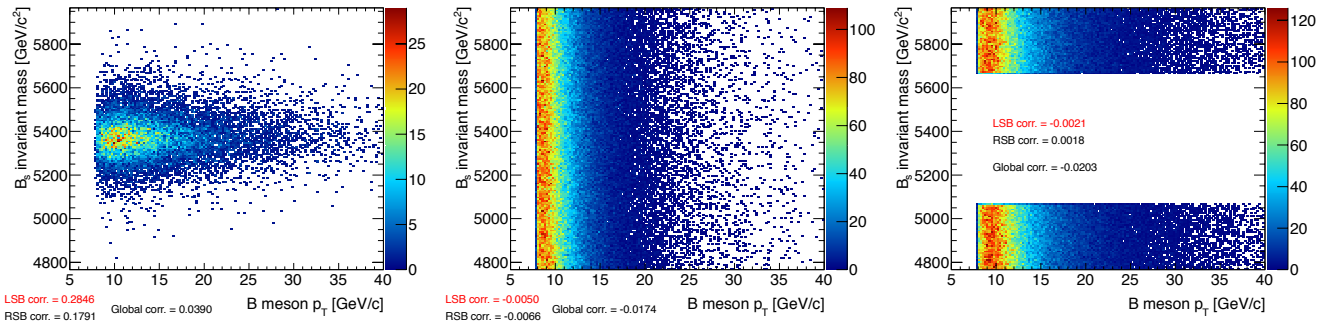
(a) d_0^{min}



(b) ZCA



(c) DCA

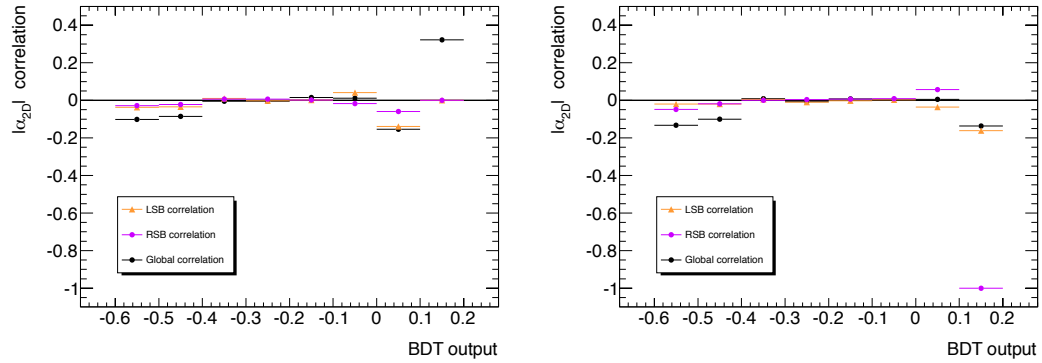


(d) B *meson* p_T

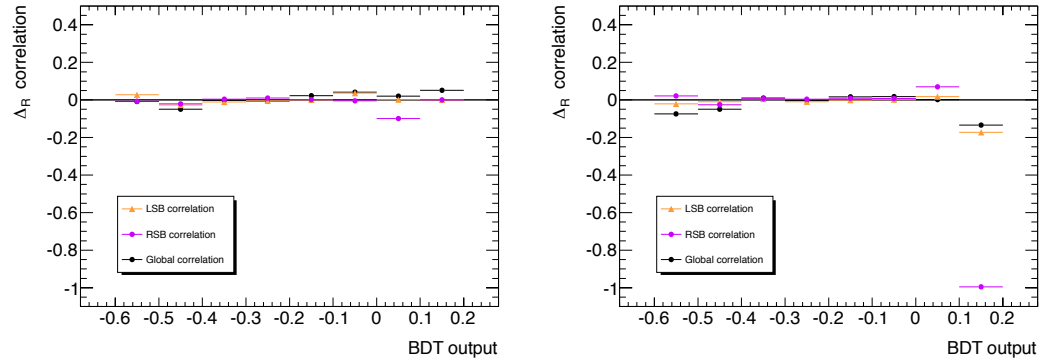
Fig. B.4. *Scatter plot (B_s mass - discriminating variables) for $B_s^0 \to \mu^+\mu^-$ signal (left), $b\bar{b} \to \mu^+\mu^- X$ background (middle) and 2011 un-blinded data (right) for the discriminating variables d_0^{min}, ZCA, DCA and B meson p_T. The global correlation, as well as the correlation on the right (RSB) and left (LSB) sidebands is reported in the plots.*

B.2 BDT Correlation of the Discriminating Variables

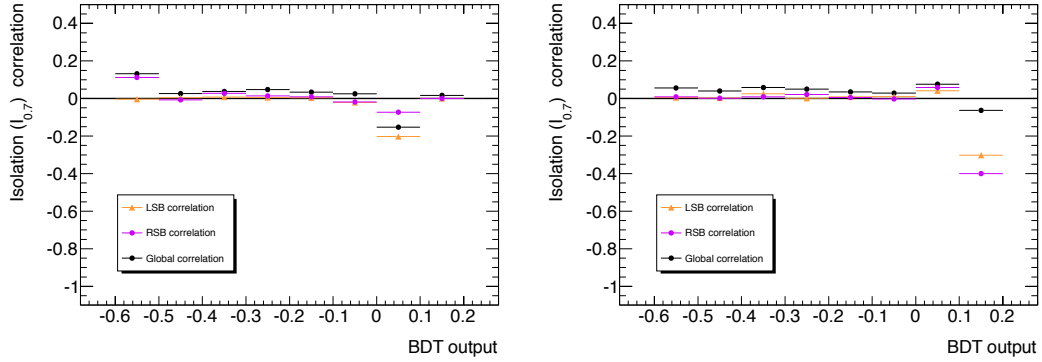
The following figures report the behaviour of the mass correlation for all the discriminating variable used in the 2011 analysis and detailed in section 5.5.1 in various BDT bins.



(a) *Pointing angle* $|\alpha_{2D}|$

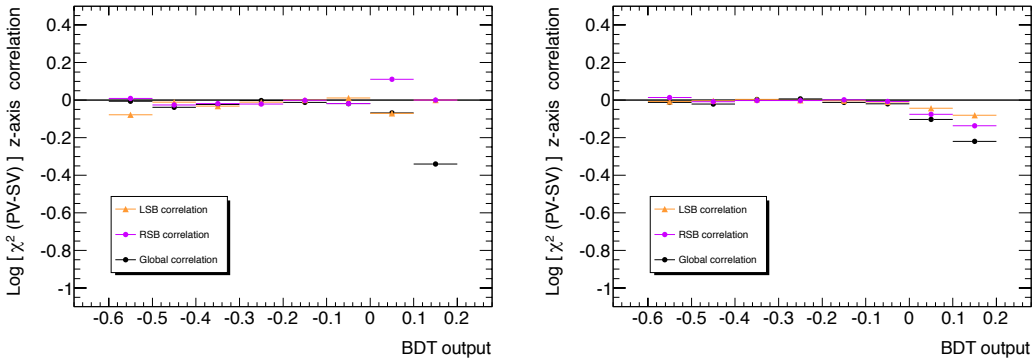


(b) ΔR

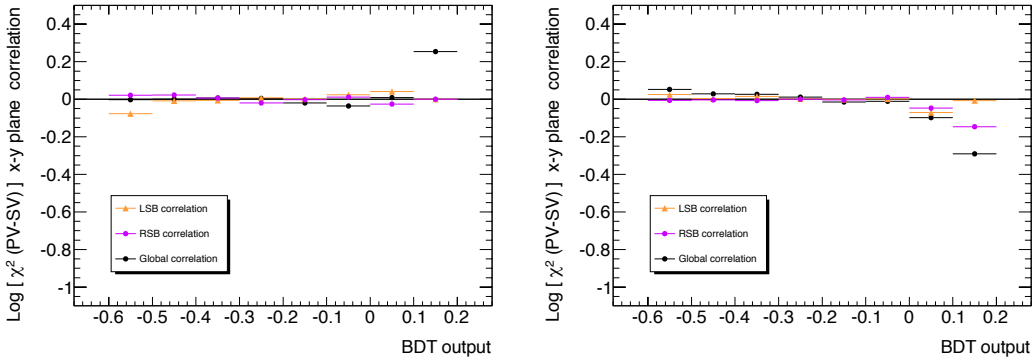


(c) *Isolation* ($I_{0.7}$)

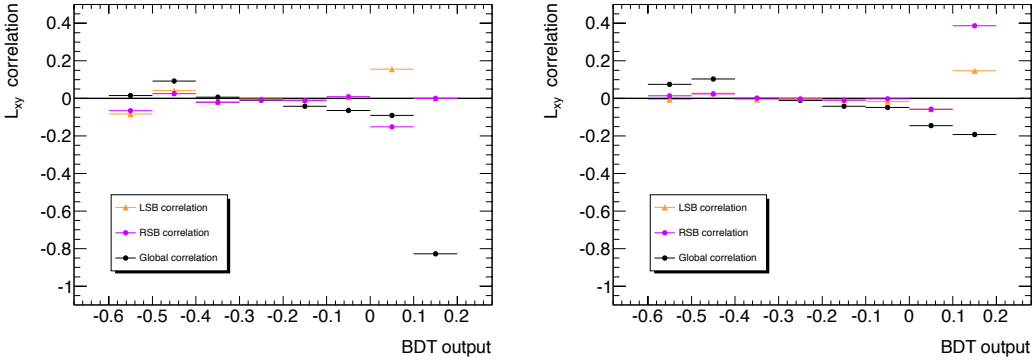
Fig. B.5. *Mass correlation in different BDT output bins for* $b\bar{b} \to \mu^+\mu^- X$ *background (left) and 2011 un-blinded data (right) for the discriminating variables* α_{2D}, ΔR *and Isolation.*



(a) $\chi^2(PV - SV)$ *along z-axis*

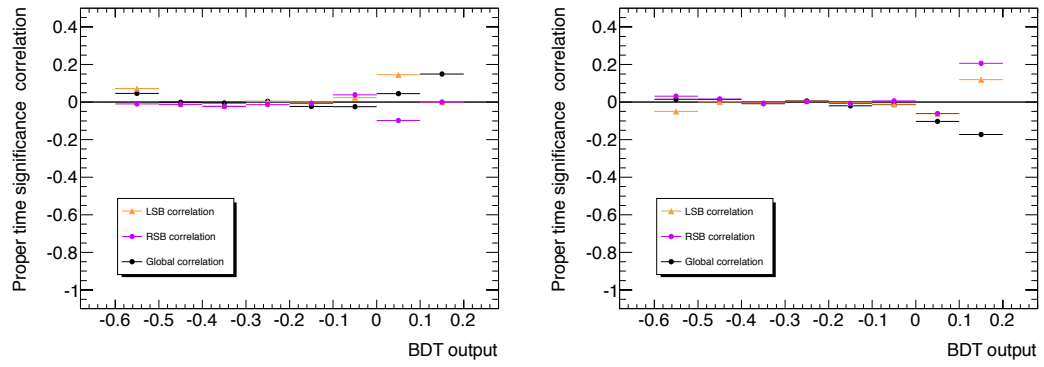


(b) $\chi^2(PV - SV)$ *in the x-y plane*

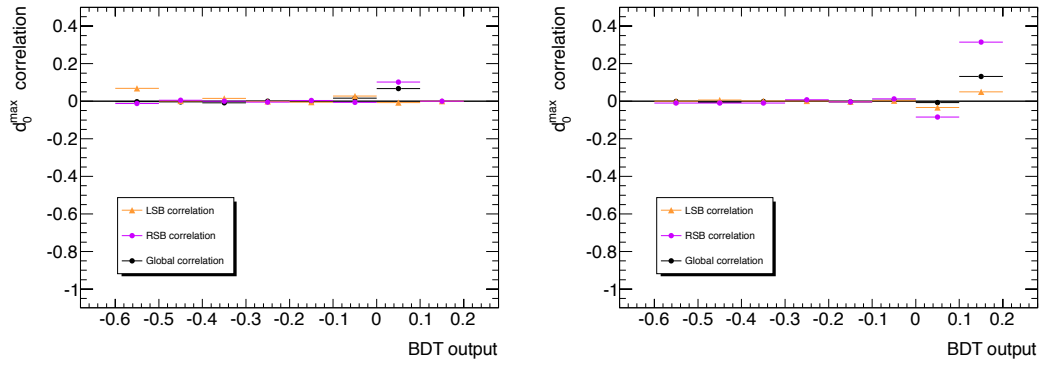


(c) L_{xy}

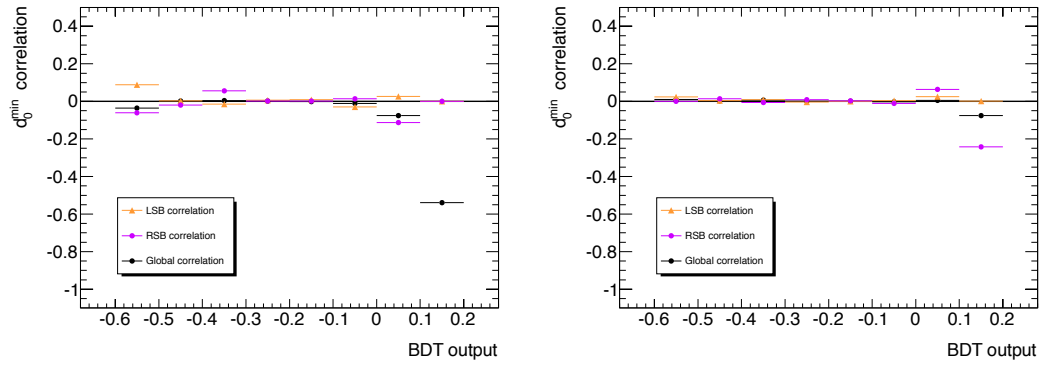
Fig. B.6. *Mass correlation in different BDT output bins for* $b\bar{b} \to \mu^+\mu^- X$ *background (left) and 2011 un-blinded data (right) for the discriminating variables* $\chi^2(PV - SV)$ *along z-axis,* $\chi^2(PV - SV)$ *in the x-y plane and* L_{xy}.



(a) *Proper time significance*

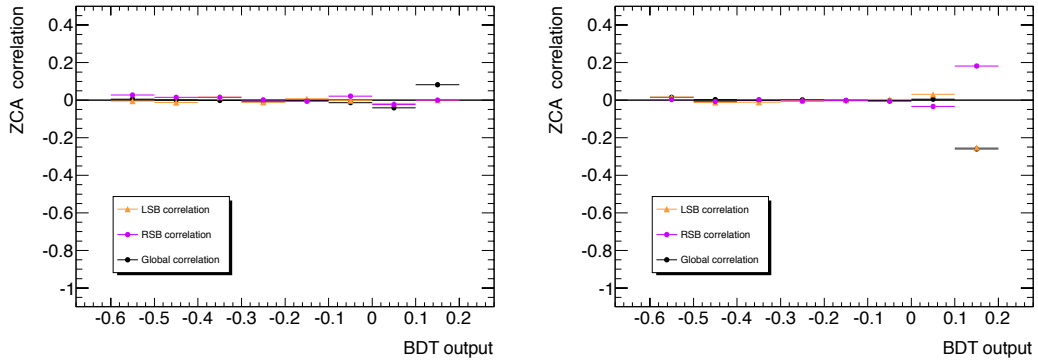


(b) d_0^{max}

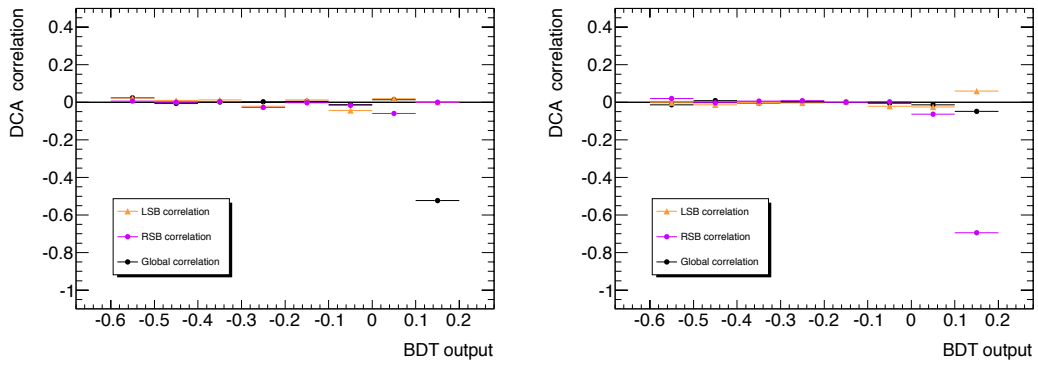


(c) d_0^{min}

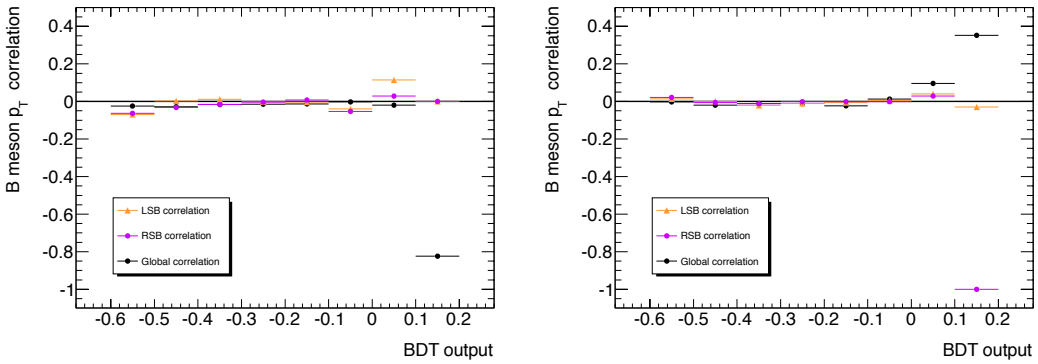
Fig. B.7. *Mass correlation in different BDT output bins for* $b\bar{b} \to \mu^+\mu^- X$ *background (left) and 2011 un-blinded data (right) for the discriminating variables proper time significance,* d_0^{max} *and* d_0^{min}.



(a) *ZCA*

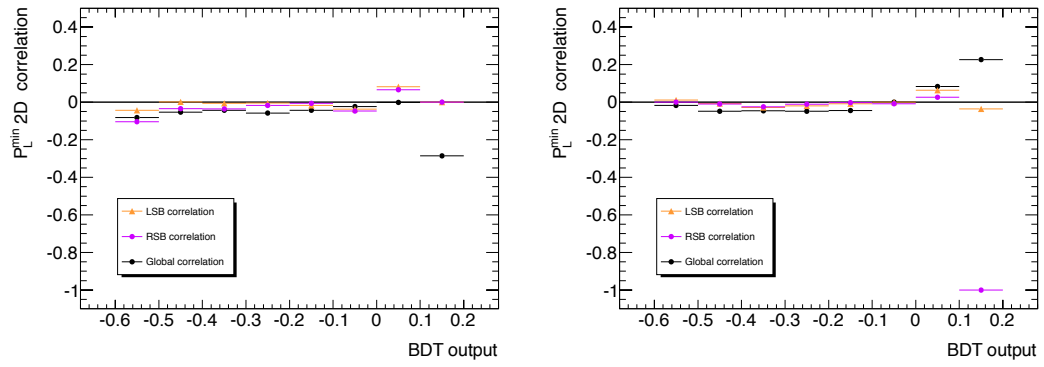


(b) *DCA*



(c) *B meson* p_T

Fig. B.8. *Mass correlation in different BDT output bins for* $b\bar{b} \to \mu^+\mu^- X$ *background (left) and 2011 un-blinded data (right) for the discriminating variables ZCA, DCA and* p_T^B.

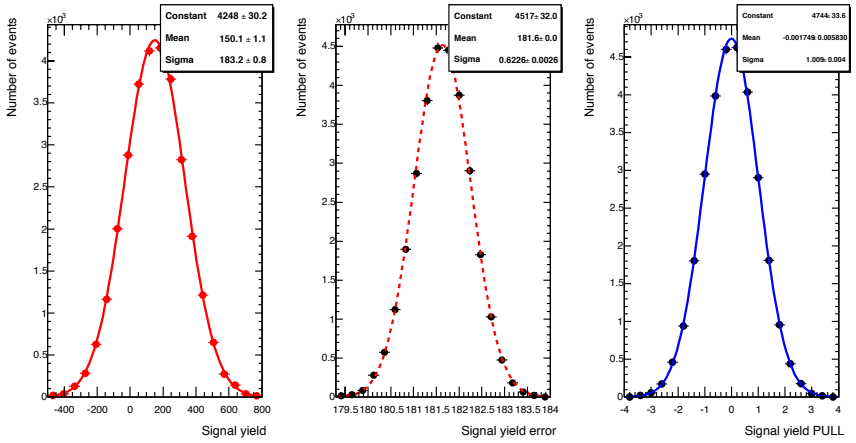


(a) p_L^{min}

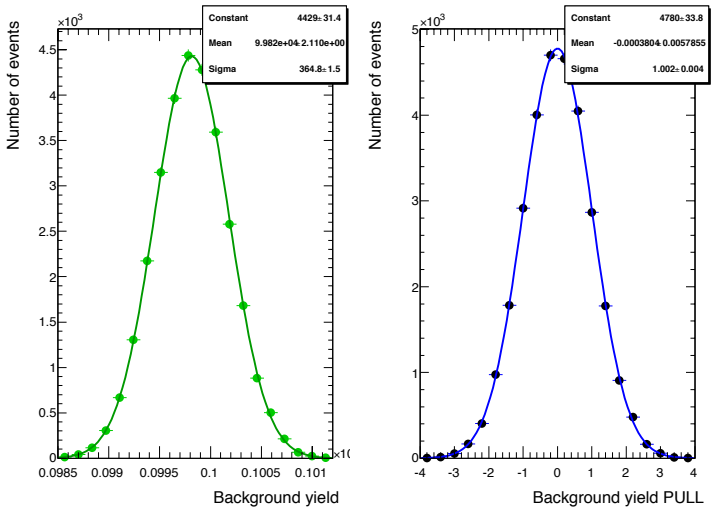
Fig. B.9. *Mass correlation in different BDT output bins for* $b\bar{b} \to \mu^+\mu^- X$ *background (left) and 2011 un-blinded data (right) for the discriminating variable* p_L^{min}.

B.3 New Fit Strategies

B.3.1 Invariant Mass Fit in Different MVA Categories

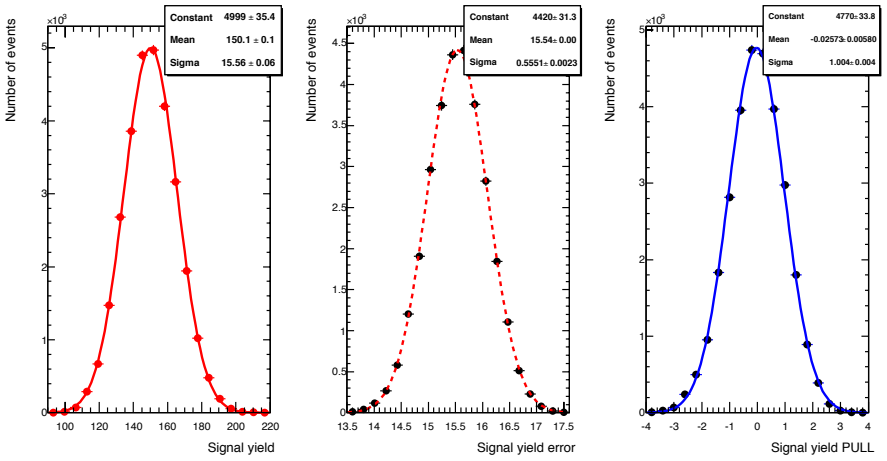


(a) *BIN 1: signal yield and pull*

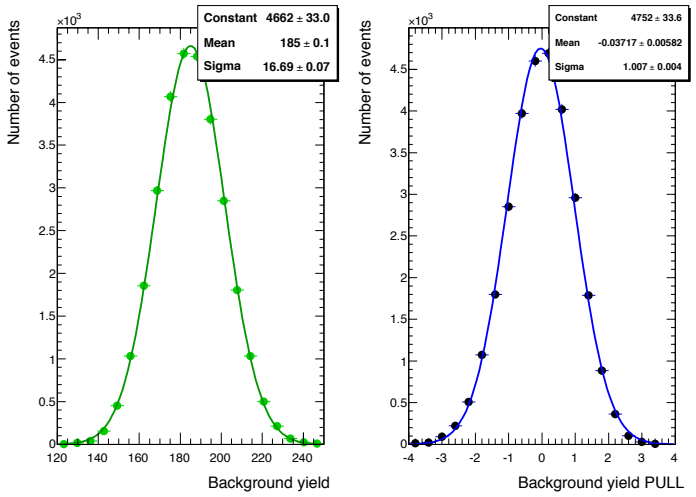


(b) *BIN 1: combinatorial background yield*

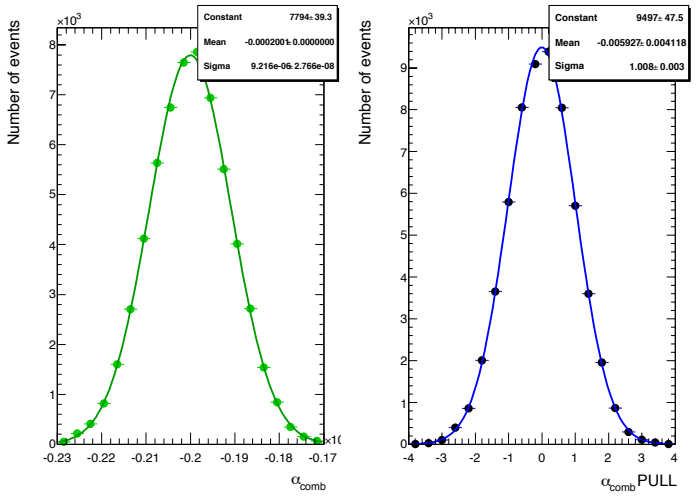
Fig. B.10. *Distribution of the parameters of interest with the related pull (obtained on 30,000 toy MC experiments) for the BDT categorised invariant mass fit obtained generating 150 signal events per bin and a 100,000 in total (99,815 and 185 events respectively in the first and second bin).*



(a) *BIN 2: signal yield and pull*



(b) *BIN 2: combinatorial background*

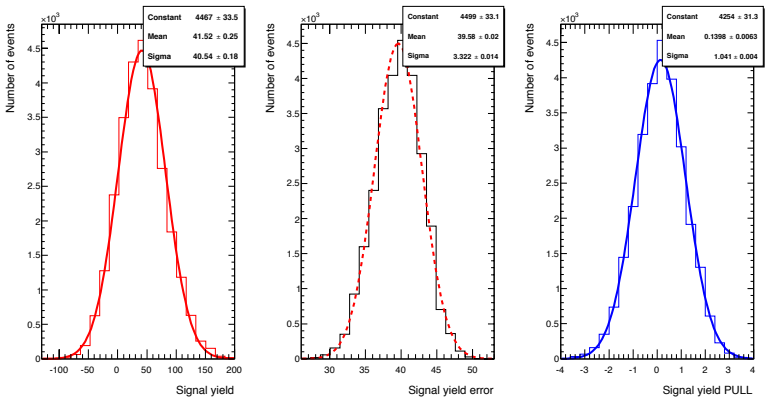


(c) *BIN 1+BIN 2: combinatorial background slope*

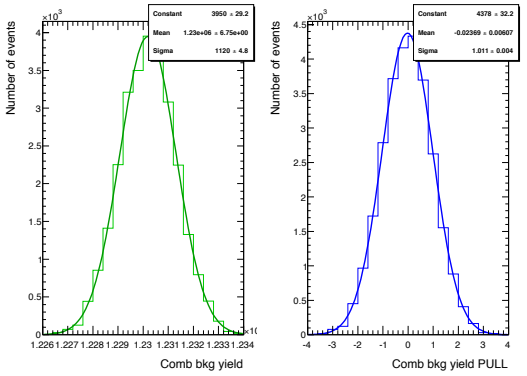
Fig. B.11. *Distribution of the parameters of interest with the related pull (obtained on 30,000 toy MC experiments) for the BDT categorised invariant mass fit obtained generating 150 signal events per bin and a 100,000 in total (99,815 and 185 events respectively in the first and second bin).*

B.3.2 Fit Tests on 2012 "setup"

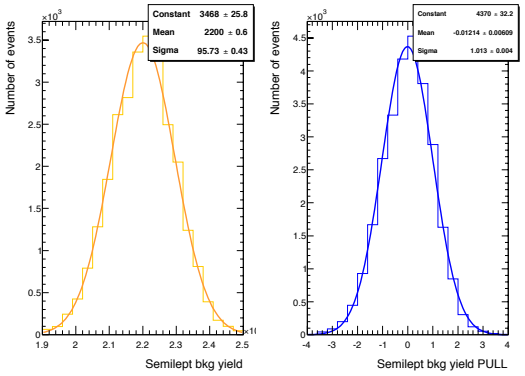
B.3.2.1 2D Fit



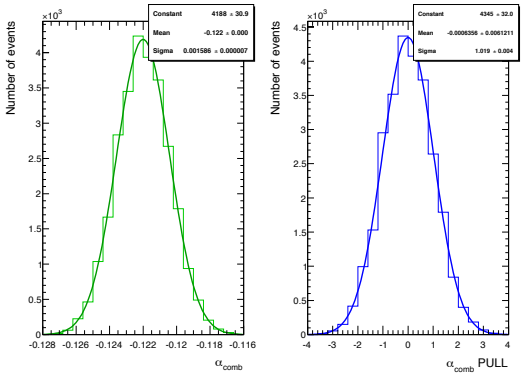
(a) *Signal yield, signal yield error and pull*



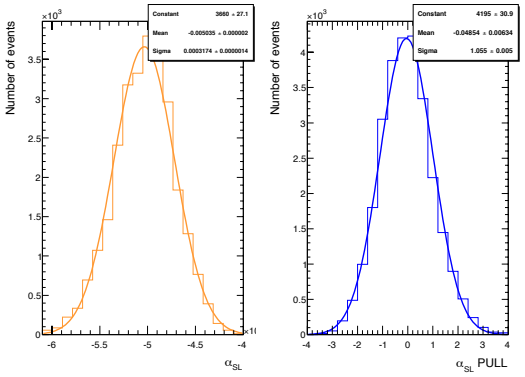
(b) *Combinatorial background yield*



(c) *Low-mass background yield*



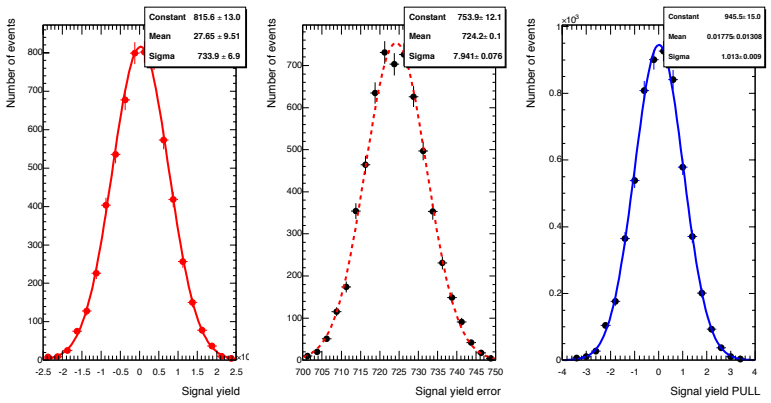
(d) *Combinatorial background shape*



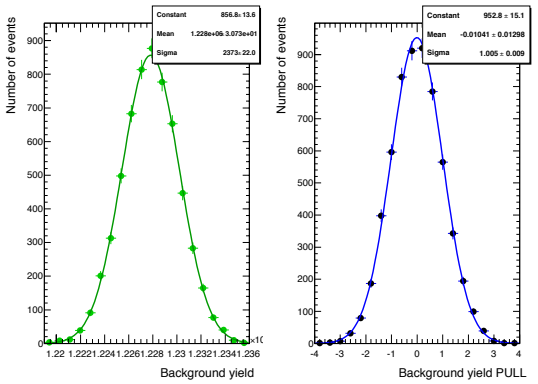
(e) *Low-mass background shape*

Fig. B.12. *Distribution of the parameters of interest with the related pull for the 2D invariant mass fit obtained running 30,000 toy MC experiments and generating the statistic expected in 2012.*

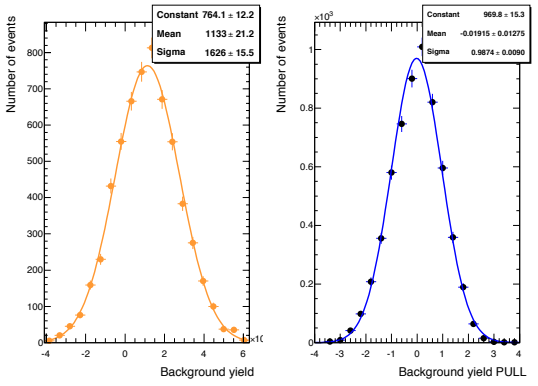
B.3.2.2 Categorised MVA Mass Fit



(a) *BIN 1: signal yield and pull*

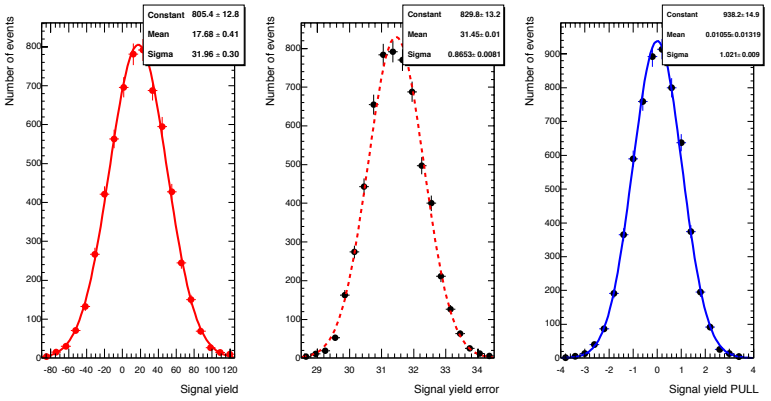


(b) *BIN 1: combinatorial background yield*

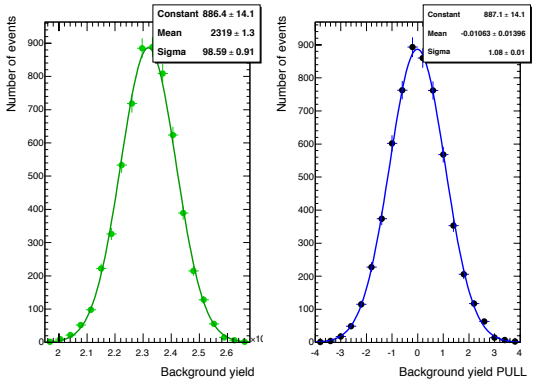


(c) *BIN 1: low-mass background*

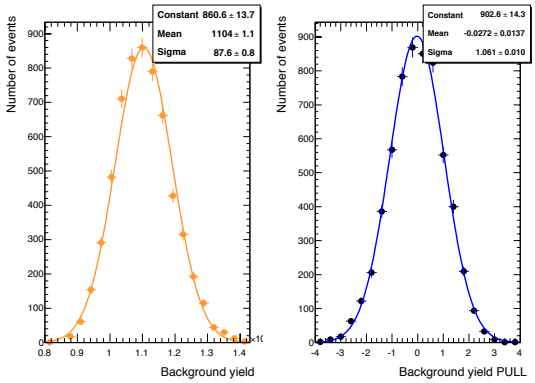
Fig. B.13. *Distribution of the parameters of interest with the related pull for the BDT categorised invariant mass fit obtained running 30,000 toy MC experiments and generating the statistic expected in 2012.*



(a) *BIN 2: signal yield and pull*

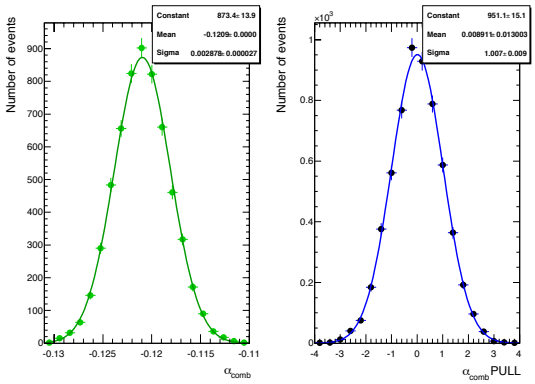


(b) *BIN 2: combinatorial background*

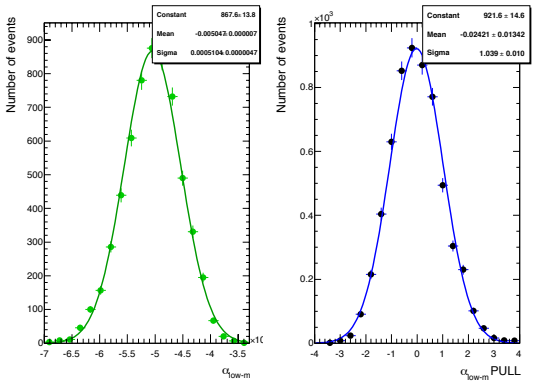


(c) *BIN 2: low-mass background*

Fig. B.14. *Distribution of the parameters of interest with the related pull for the BDT categorised invariant mass fit obtained running 30,000 toy MC experiments and generating the statistic expected in 2012.*



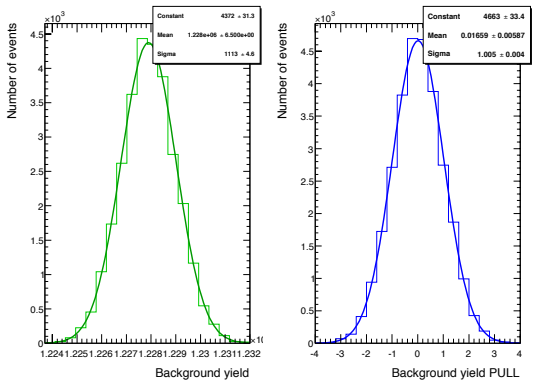
(a) *BIN 1 + BIN 2: combinatorial background slope*



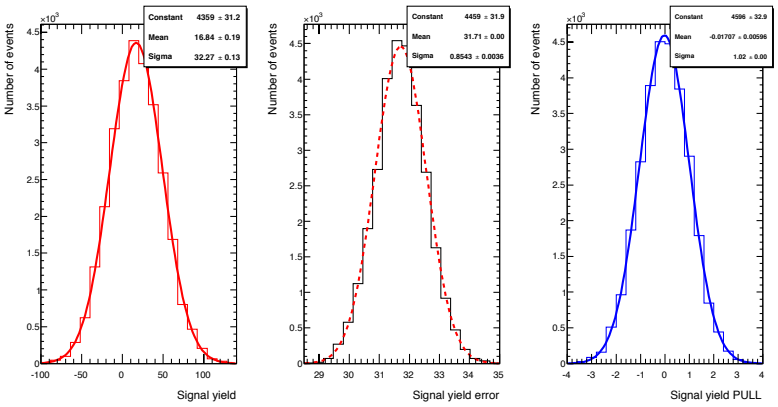
(b) *BIN 1 + BIN 2: low-mass background slope*

Fig. B.15. *Distribution of the parameters of interest with the related pull for the BDT categorised invariant mass fit obtained running 30,000 toy MC experiments and generating the statistic expected in 2012.*

B.3.2.3 Categorised MVA Mass Fit (B_s and semi-leptonic yields fixed in bin 1)

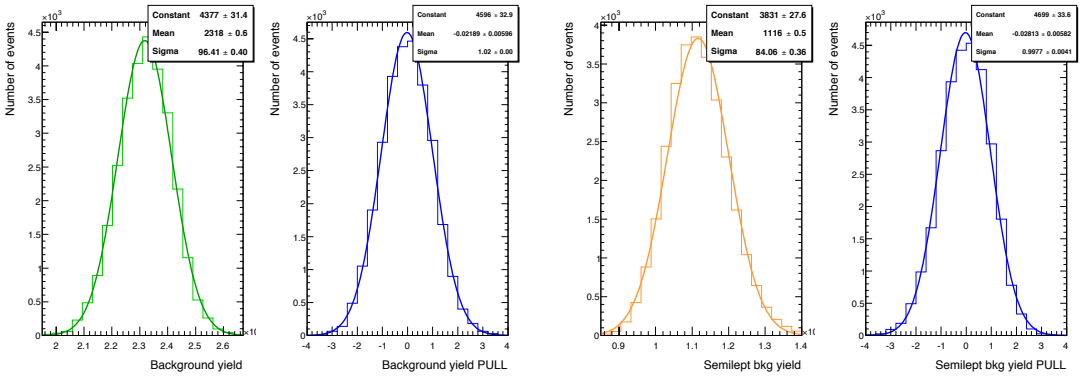


(a) *BIN 1: combinatorial background yield*

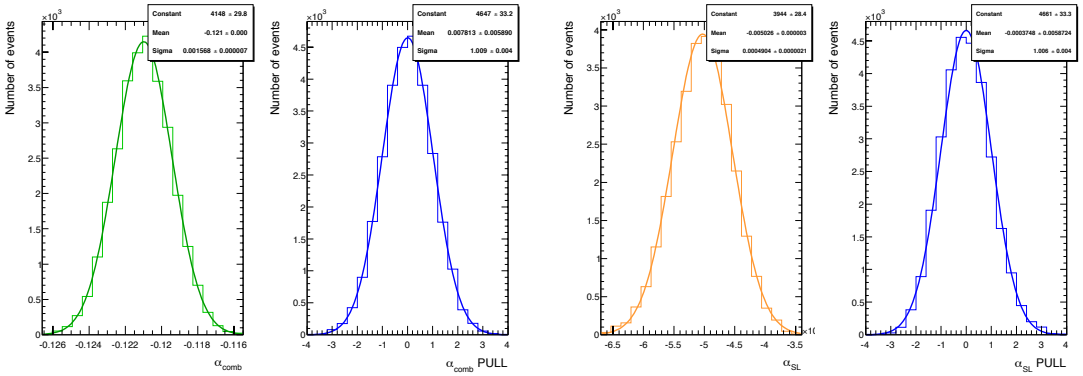


(b) *BIN 2: signal yield and pull*

Fig. B.16. *Distribution of the parameters of interest with the related pull for the BDT categorised invariant mass fit obtained running 30,000 toy MC experiments and generating the statistic expected in 2012 (B_s and low-mass yields are fixed in the first bin).*



(a) *BIN 2: combinatorial background* **(b)** *BIN 2: low-mass background*



(c) *BIN 1 + BIN 2: combinatorial background slope* **(d)** *BIN 1 + BIN 2: low-mass background slope*

Fig. B.17. *Distribution of the parameters of interest with the related pull for the BDT categorised invariant mass fit obtained running 30,000 toy MC experiments and generating the statistic expected in 2012 (B_s and low-mass yields are fixed in the first bin).*

Bibliography

[1] C. Balázs, T. Li and J. L. Newstead, *Thermal dark matter implies new physics not far above the weak scale*, JHEP 08 (2014) 061, arXiv:1403.5829 [hep-ph].

[2] M. Alam et al. (CLEO Collaboration), *First Measurement of the Rate for the Inclusive Radiative Penguin Decay* $b \to s\,\gamma$. Phys. Rev. Lett. **74** (1995) 2885-2889.

[3] C. Bobeth et al. , $B_{s,d} \to \ell^+\ell^-$ *in the Standard Model with Reduced Theoretical Uncertainty*, Phys. Rev. Lett. **112** (2014) 101801, arXiv:1311.0903 [hep-ex].

[4] ATLAS Collaboration, *Limit on* $B_s^0 \to \mu^+\mu^-$ *branching fraction based on 4.9 fb−1 of integrated luminosity*, ATLAS-CONF-2013-076, `http://cds.cern.ch/record/1562934/files/ATLAS-CONF-2013-076.pdf`.

[5] LHCb Collaboration, *Measurement of the* $B_s^0 \to \mu^+\mu^-$ *Branching Fraction and Search for* $B^0 \to \mu^+\mu^-$ *with the LHCb Experiment*, Phys. Rev. Lett. **111** (2013) 101805, arXiv:1307.5024 [hep-ex].

[6] CMS Collaboration, *Measurement of the* $B_s^0 \to \mu^+\mu^-$ *Branching Fraction and Search for* $B^0 \to \mu^+\mu^-$ *with the CMS Experiment*, Phys. Rev. Lett. **111** (2013) 101804.

[7] CMS Collaboration, *Measurement of the* $B_s^0 \to \mu^+\mu^-$ *Branching Fraction and Search for* $B^0 \to \mu^+\mu^-$ *with the CMS Experiment*, arXiv:1307.5025 [hep-ex].

[8] ATLAS Collaboration, *Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC*, Phys. Lett. B **716** (2012) 1-29, arXiv:1207.7214 [hep-ex].

[9] CMS Collaboration, *Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC*, Phys. Lett. B **716** (2012) 30-61, arXiv:1207.7235 [hep-ex].

[10] L. Accardo et al. (AMS Collaboration), *High Statistics Measurement of the Positron Fraction in Primary Cosmic Rays of 0.5–500 GeV with the Alpha Magnetic Spectrometer on the International Space Station*, Phys. Rev. Lett. **113** (2014) 121101.

[11] G. Isidori, *Flavor physics and CP violation*, arXiv:1302.0661 [hep-ph].

[12] T. E. Browder et al., *New physics at a Super Flavor Factory*, Rev. Mod. Phys. **81** (2009) 1887, arXiv:0802.3201 [hep-ph].

[13] R. D. Peccei and H. R. Quinn, *CP Conservation in the Presence of Pseudoparticles*, Phys. Rev. Lett. **38** (1977) 1440.

[14] M. Antonelli et al., *Flavor Physics in the Quark Sector*, arXiv:0907.5386 [hep-ex].

[15] Particle Data Group, *The CKM quark-mixing matrix*. `http://pdg.lbl.gov/2014/reviews/rpp2014-rev-ckm-matrix.pdf`

[16] M. Bona et al. (UTfit Collaboration), *The Unitarity Triangle Fit in the Standard Model and Hadronic Parameters from Lattice QCD: A Reappraisal after the Measurements of Δm_s and BR($B \to \tau\nu_\tau$)*, JHEP 10 (2006) 081, hep-ph/0606167. **Online update at:** `http://www.utfit.org/UTfit/ResultsSummer2014PostMoriond`

[17] CKMfitter Collaboration. **Online update at:** `http://ckmfitter.in2p3.fr/www/html/ckm_main.html`

[18] S. Weinberg, *Baryon And Lepton Nonconserving Processes*, Phys. Rev. Lett. **43** (1979) 1566.

[19] M. Artuso et al., *B, D and K decays*, Eur. Phys. J. C **57** (2008) 309, arXiv:0801.1833 [hep-ex].

[20] S. L. Glashow, J. Iliopoulos, and L. Maiani, *Weak Interactions with Lepton-Hadron Symmetry*, Phys. Rev. D **2** (1970) 1285 - 1292.

[21] K. G. Wilson, *Nonlagrangian models of current algebra*, Phys. Rev. **179** (1969) 1499-1512.

[22] K. G. Wilson, *The Renormalization Group and Strong Interactions*, Phys. Rev. D **3** (1971) 1818.

[23] E. Witten, *Short Distance Analysis of Weak Interactions*, Nucl. Phys. B **122** (1977) 109.

[24] F. J. Gilman, M. B. Wise, *Effective Hamiltonian for $\Delta S = 1$ Weak Nonleptonic Decays in the Six Quark Model*, Phys. Rev. D **20** (1979) 2392.

[25] K. G. Wilson and W. Zimmermann, *Operator product expansions and composite field operators in the general framework of quantum field theory*, Comm. Math. Phys. **24** (1972) 87-106.

[26] K. De Bruyn et al., *Branching Ratio Measurements of B_s Decays*, Phys. Rev. D **86**, 014027 (2012), arXiv:1204.1735 [hep-ph].

[27] K. De Bruyn et al. , *Probing New Physics via the $B_s^0 \to \mu^+\mu^-$ Effective Lifetime*, Phys. Rev. Lett. **109** (2012) 041801, arXiv:1204.1737 [hep-ex].

[28] The CMS and LHCb Collaborations, *Observation of the rare $B_s^0 \to \mu^+\mu^-$ decay from the combined analysis of CMS and LHCb data*, arXiv:1411.4413 [hep-ex].

[29] Y. Amhis et al. (Heavy Flavor Averaging Group), *Averages of b-hadron, c-hadron, and tau-lepton properties as of early 2012*, arXiv:1207.1158 [hep-ex]; Y. Amhis et al. (Heavy Flavor Averaging Group), updates at `http://www.slac.stanford.edu/xorg/hfag`.

[30] A. Lenz and U. Nierste, *Numerical updates of lifetimes and mixing parameters of B mesons*, arXiv:1102.4274 [hep-ph].

[31] , G. Buchalla, A. J. Buras, M. E. Lautenbacher, *Weak Decays Beyond Leading Logarithms*, Rev. Mod. Phys. **68** (1996) 1125-1144, arXiv:hep-ph/9512380.

[32] S. Aoki et al., *Review of lattice results concerning low energy particle physics*, arXiv:1310.8555 (2013) [updates at `http://itpwiki.unibe.ch/`].

[33] Particle Data Group, J. Beringer et al., *Review of particle physics*, Phys. Rev. D **86** (2012) 010001, and 2013 partial update for the 2014 edition.

[34] Heavy Flavor Averaging Group, Y. Amhis et al., *Averages of b-hadron, c-hadron, and τ-lepton properties as of early 2012*. arXiv:1207.1158 (2012) [updated results and plots available at: `http://www.slac.stanford.edu/xorg/hfag/`].

[35] HFAG Collaboration, *b-hadron fractions at high energy (averages for the PDG 2014 review)*, **online update at:** `http://www.slac.stanford.edu/xorg/hfag/osc/PDG_2014/#FRAC`.

[36] W. Altmannshofer and D. Straub, *Cornering New Physics in $b \to s$ Transitions*, FERMILAB-PUB-12-257-T, arXiv:1206.0273v3.

[37] G. Isidori, Y. Nir, G. Perez, *Flavor Physics Constraints for Physics Beyond the Standard Model*, Annu. Rev. Nucl. Part. Sci. **60** (2010) 355-380, arXiv:1002.0900 [hep-ph].

[38] G. DÕAmbrosio, G. F. Giudice, G. Isidori and A. Strumia, *Minimal Flavour Violation: an effective field theory approach*, Nucl. Phys. B **645** (2002) 155, hep-ph/0207036.

[39] M. Bona et al., *Model-independent constraints on $\Delta F = 2$ operators and the scale of New Physics*, JHEP **0803** (2008) 049, arXiv:0707.0636 [hep-ph]. **online update at:** `http://www.utfit.org/UTfit/ResultsSummer2014PostMoriond`.

[40] LHCb Collaboration, *Letter of Intent for the LHCb Upgrade*, CERN, March, 2011, Tech. Rep., CERN-LHCC-2011-001.

[41] S. P. Martin, *A Supersymmetry Primer*, arXiv:hep-ph/9709356.

[42] G. L. Kane, *Study of Constrained Minimal Supersymmetry*, Phys. Rev. D **49** (1994) 6173-6210, arXiv:hep-ph/9312272.

[43] M. Drees, M. M. Nojiri, *The Neutralino Relic Density in Minimal N=1 Supergravity*, Phys. Rev. D **47** (1993) 376-408, arXiv:hep-ph/9207234.

[44] H. Baer, M. Brhlik, *Cosmological Relic Density from Minimal Supergravity with Implications for Collider Physics*, Phys. Rev. D **53** (1996) 597-605, arXiv:hep-ph/9508321.

[45] O. Buchmueller et al., *The CMSSM and NUHM1 in Light of 7 TeV LHC, $B_s^0 \to \mu^+\mu^-$ and XENON100 Data*, arXiv:1207.7315 [hep-ph].

[46] S. S. AbdusSalam et al., *Benchmark Models, Planes, Lines and Points for Future SUSY Searches at the LHC*, Eur. Phys. J. C **71** (2011) 1835, arXiv:1109.3859 [hep-ph].

[47] H. Baer et al., *Neutralino cold dark matter in a one-parameter extension of the minimal supergravity model*, Phys. Rev. D **71** (2005) 095008, arXiv:hep-ph/0412059;

[48] M. Blanke, A. J. Buras, B. Duling, K. Gemmler and S. Gori, *Rare K and B Decays in a Warped Extra Dimension with Custodial Protection*, JHEP **0903** (2009) 108, arXiv:0812.3803 [hep-ph]

[49] A. J. Buras, B. Duling, T. Feldmann, T. Heidsieck, C. Promberger and S. Recksiegel, *Patterns of Flavour Violation in the Presence of a Fourth Generation of Quarks and Leptons*, JHEP **1009** (2010) 106, arXiv:1002.2126 [hep-ph].

[50] D. M. Straub, *Overview of constraints on new physics in rare B decays*, Eur. Phys. J. C **72** (2012) 2243, arXiv:1205.6094 [hep-ph].

[51] K. Agashe and C. D. Carone, *Supersymmetric Flavor Models and the $B \to \phi K_S$ Anomaly*, Phys. Rev. **D** 68 (2003) 035017, hep-ph/0304229.

[52] G. G. Ross, L. Velasco-Sevilla and O. Vives, *Spontaneous CP violation and Non-Abelian Family Symmetry in SUSY*, Nucl. Phys. **B** 692 (2004) 50, hep-ph/0401064.

[53] S. Antusch, S. F. King and M. Malinsky, *Solving the SUSY Flavour and CP Problems with SU(3) Family Symmetry*, JHEP **0806** (2008) 068, arXiv:0708.1282 [hep-ph].

[54] L. J. Hall and H. Murayama, *A Geometry of the Generations*, Phys. Rev. Lett. **75** (1995) 3985, hep-ph/9508296.

[55] W. Altmannshofer, A. J. Buras, S. Gori, P. Paradisi and D. M. Straub, *Anatomy and Phenomenology of FCNC and CPV Effects in SUSY Theories*, Nucl. Phys. B **830** (2010) 17, arXiv:0909.1333 [hep-ph].

[56] L. Evans and P. Bryant (editors), *LHC machine*, JINST **3** (2008) S088001.

[57] CERN, *Interim summary report on the analysis of the 19 September 2008 incident at the LHC*, EDMS 973073, https://edms.cern.ch/file/973073/1/Report_on_080919_incident_at_LHC__2_.pdf.

[58] https://espace.cern.ch/be-dep/BEDepartmentalDocuments/BE/LHC_Schedule_2015.pdf.

[59] ATLAS Collaboration, *The ATLAS Experiment at the CERN Large Hadron Collider*, JINST **3** (2008) S08003.

[60] O. Brüning et al. , *LHC Design Report*, CERN-2004-003-V-1, http://cds.cern.ch/record/782076?ln=en.

[61] CMS Collaboration, *The CMS experiment at the CERN LHC*, JINST **3** (2008) S08004.

[62] TOTEM Collaboration, *The TOTEM Experiment at the CERN Large Hadron Collider*, JINST **3** (2008) S08007.

[63] ALICE Collaboration, *The ALICE experiment at the CERN LHC*, JINST **3** (2008) S08002.

[64] LHCb Collaboration, *The LHCb Detector at the LHC*, JINST **3** (2008) S08005.

[65] CERN, *LEP design report* Volume 2 *The LEP main ring*, CERN-LEP-84-01 (1984).

[66] J.P. Blewett, *200 GeV intersecting storage accelerators*, Proceedings of the 8^{th} International Conference on High-Energy Accelerators, CERN, Geneva Switzerland (1971), http://lss.fnal.gov/conf/C710920/p501.pdf.

[67] TeVI Group *Design report Tevatron 1 Project*, FERMILAB-DESIGN-1982-01, http://lss.fnal.gov/archive/design/fermilab-design-1982-01.pdf.

[68] TeVI Group *Design report Tevatron 1 Project*, FERMILAB-DESIGN-1983-01, http://lss.fnal.gov/archive/design/fermilab-design-1983-01.pdf.

[69] TeVI Group, *Design report Tevatron 1 Project*, FERMILAB-DESIGN-1984-01, http://lss.fnal.gov/archive/design/fermilab-design-1984-01.pdf.

[70] Gupta, R., Ramberger, S., Russenschuck, S. *Field quality optimization in a common coil magnet design*, 16th International Conference on Magnet Technology, MT16, Fl, USA, 1999.

[71] H. Cuna, J. Contreras and F. Zimmermann, *Simulations of electron-cloud heat load for the cold arcs of the CERN Large Hadron Collider and its high-luminosity upgrade scenarios*, Phys. Rev. ST Accel. Beams **15** (2012) 051001 .

[72] A. Yamamoto et al., *The ATLAS central solenoid*, Nucl. Instrum. Meth. **A 584** (2008) 53.

[73] A. Yamamoto et al., *ATLAS Superconducting Central Solenoid*, `http://atlas.kek.jp/sub/poster/index.html` (poster).

[74] ATLAS Collaboration, *The ATLAS Inner Detector commissioning and calibration*, Eur. Phys. J. **C70** (2010) 787 - 821.

[75] ATLAS Collaboration, *Expected Performance of the ATLAS Experiment - Detector, Trigger and Physics*, arXiv:0901.0512 [hep-ex].

[76] ATLAS Collaboration, *ATLAS pixel detector electronics and sensors*, JINST **3** (2008) P07007.

[77] ATLAS Collaboration, *Operation and performance of the ATLAS semiconductor tracker*, arXiv:1404.7473 [hep-ex].

[78] J.R.Pater, *The ATLAS SemiConductor Tracker Operation and Performance*, JINST **7** (2012) C04001.

[79] *F. Campabadal et al.*, *Design and performance of the ABCD3TA ASIC for readout of silicon strip detectors in the ATLAS semiconductor tracker*, Nucl. Instrum. Meth. **A 552** (2005) 292.

[80] ATLAS TRT Collaboration, *The ATLAS Transition Radiation Tracker (TRT) proportional drift tube: design and performance*, JINST **3** (2008) P02013.

[81] ATLAS collaboration, *Liquid argon calorimeter technical design report*, CERN-LHCC-96-041, `http://cdsweb.cern.ch/record/331061`.

[82] ATLAS collaboration, *Tile calorimeter technical design report*, CERN-LHCC-96-042, `http://cdsweb.cern.ch/record/331062`.

[83] ATLAS Collaboration, *ATLAS muon spectrometer: Technical design report*, CERN-LHCC-97-22, ATLAS-TDR-10, `http://atlas.web.cern.ch/Atlas/GROUPS/MUON/TDR/pdf_final/mTDR.pdf`.

[84] *https://cds.cern.ch/record/1631701.*

[85] *https://twiki.cern.ch/twiki/bin/view/AtlasPublic/MuonPerformancePublicPlotsSimulation.*

[86] ATLAS Collaboration, *Luminosity public results*, `https://twiki.cern.ch/twiki/bin/view/AtlasPublic/LuminosityPublicResults`

[87] ATLAS Collaboration, *Improved luminosity determination in pp collisions at* $\sqrt{s} = 7$ *TeV using the ATLAS detector at the LHC*, Eur. Phys. J. C **73** (2013) 2518.

[88] J. Zhu, *The RPC-based proposal for the ATLAS forward muon trigger upgrade in view of super-LHC*, Proceedings Of Science (PoS) (RPC2012) 062, http://pos.sissa.it/archive/conferences/159/062/RPC2012_062.pdf.

[89] ATLAS Collaboration, B-physics trigger public results, https://twiki.cern.ch/twiki/bin/view/AtlasPublic/BPhysicsTriggerPublicResults.

[90] ATLAS Collaboration, *Measurement of Upsilon production in 7 TeV pp collisions at ATLAS*, Phys. Rev. **D87** (2013) 052004, arXiv:1211.7255. More details can be found here: http://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PAPERS/BPHY-2011-06/.

[91] I. Gorelov et al., *A measurement of Lorentz angle and spatial resolution of radiation hard silicon pixel sensors*, Nuclear Instruments and Methods in Physics Research Section A **481** (2002), 204-221.

[92] V. Bartsch et al., *An algorithm for calculating the Lorentz angle in silicon detectors*, Nucl. Instr. Meth. A **497** (2003) 389-396.

[93] C. Jacoboni, C. Canali, G. Ottaviani and A. Alberighi Quaranta, *A review of some charge transport properties of silicon* , Solid State Electronics **20** (1977) 77 - 89.

[94] C. Canali, G. Majini, R. Minder and G. Ottaviani, *Electron and hole drift velocity measurements in silicon and their empirical relation to electric field and temperature*, IEEE Trans. Electron Devices **ED-7** (1970) 481.

[95] J. Becker, E. Fretwurst , R. Klanner, *Measurements of charge carrier mobilities and drift velocity saturation in bulk silicon of <111> and <100> crystal orientation at high electric fields*, Solid-State Electronics **56** (2011) , 104 - 110.

[96] M. Bona, M. Castanheira, S. McMahon, E. Piccaro, P. Ward, *Lorentz Angle Analysis and Measurement for the ATLAS SCT*, ATL-COM-INDET-2011-114.

[97] ROOT Data Analysis Framework, http://root.cern.ch/root/html/TProfile.html.

[98] The CMS Collaboration, *Commissioning and Performance of the CMS Silicon Strip Tracker with Cosmic Ray Muons*, CMS PAPER CFT-09-002.

[99] T. Kohriki et al., *Evaluation of Contact Resistance of Silver-Loaded Epoxy with Aluminized Backplane of Silicon Microstrip Sensors*, NSS '07. IEEE **1** 673-676.

[100] Phillips, P W, *ATLAS SCT power supply system*, Topical Workshop on Electronics for Particle Physics, Prague, Czech Republic, 03 - 07 Sep 2007, pp.365-368, https://cds.cern.ch/record/1091485?ln=en

[101] F. Campabadal et al., *Design and performance of the ABCD3TA ASIC for readout of silicon strip detectors in the ATLAS semiconductor tracker*, Nucl. Instrum. Meth. A **552** (2005) 292-328.

[102] Model 2410-C High-Voltage SourceMeter, `http://www.keithley.com/products/dcac/voltagesource/highvoltage/?mn=2410-C`.

[103] 34970A Data Acquisition / Data Logger Switch Unit, `http://www.keysight.com/en/pd-1000001313:epsg:pro-pn-34970A/data-acquisition-data-logger-switch-unit?&cc=GB&lc=eng`.

[104] NI GPIB-USB-HS, `http://sine.ni.com/nips/cds/view/p/lang/en/nid/201586`.

[105] LabVIEW System Design Software, `http://www.ni.com/labview/`.

[106] ATLAS Collaboration, *Limit on $B_s^0 \to \mu^+\mu^-$ branching fraction based on 4.9 fb^{-1} of integrated luminosity* , ATLAS-CONF-2013-076, `http://cds.cern.ch/record/1562934/files/ATLAS-CONF-2013-076.pdf`.

[107] ATLAS Collaboration, *Search for the decay $B_s^0 \to \mu^+\mu^-$ with the ATLAS detector*, Phys. Lett. B **713** (2012) 180-196, arXiv:1204.0735 [hep-ex].

[108] B. Aubert et al. (*BABAR* Collaboration), *A Study of Time-Dependent CP-Violating Asymmetries and Flavor Oscillations in Neutral B Decays at the $\Upsilon(4S)$*, Phys. Rev. D **66** (2002) 032003, hep-ex/0201020.

[109] K. Abe et al. (BELLE Collaboration), *Observation of Mixing-induced CP Violation in the Neutral B Meson System*,Phys. Rev. D. **66** (2002) 032007, hep-ex/0202027.

[110] A. Alavi-Harati et al., *Observation of Direct CP Violation in $K_{S,L} \to \pi\pi$ Decays*, Phys. Rev. Lett. **83** (1999) 22.

[111] E. Altala et al. (Fermilab E791 Collaboration), *Search for the Pentaquark via the $P_{\bar{c}s}^0 \to \varphi\,\pi\,p$ Decay*, Phys. Rev. Lett. **81** (1998) 44;

[112] K. Arlsaka et al., *Improved upper limit on the branching ratio $B(K_L^0\,\mu^\pm e^\mp)$*, Phys. Rev. Lett. **70** (1993) 1049.

[113] ATLAS Collaboration, *Extraction of the prompt muon component in inclusive muons produced at $\sqrt{s} = 7$ TeV*, ATLAS-CONF-2010-075, `http://cds.cern.ch/record/1281366/files/ATLAS-CONF-2010-075.pdf`.

[114] T. Sjöstrand, S. Mrenna and P. Skands, *PYTHIA 6.4 Physics and Manual*, JHEP **05** (2006) 026, arXiv:hep-ph/0603175 [hep-ex].

[115] ATLAS Collaboration, *ATLAS tunes of PYTHIA6 and Pythia8 for MC11*, ATL-PHYS-PUB-2011-009, `http://cds.cern.ch/record/1363300/files/ATL-PHYS-PUB-2011-009.pdf`.

[116] S. Agostinelli et al, *Geant4 - a simulation toolkit*, Nucl. Instr. Meth. Phys. Res. A **506** 250-303.

[117] W. Lukas, *Fast Simulation for ATLAS: Atlfast-II and ISF*, Journal of Physics: Conference Series **396** (2012) 022031.

[118] A. Hoecker, P. Speckmayer, J. Stelzer, J. Therhaag, E. von Toerne, and H. Voss, *TMVA, the Toolkit for Multivariate Data Analysis with ROOT*, PoS ACAT **040** (2007), arXiv:physics/0703039 [physics.data-an].

[119] G. Punzi, *Sensitivity of searches for new signals and its optimization*, arXiv:physics/0308063v2 [physics.data-an].

[120] K. S. Cranmer, *Kernel Estimation in High-Energy Physics*, Comput.Phys.Commun. **136** (2001) 198-207, arXiv:hep-ex/0011057.

[121] Particle Data Group Collaboration, J. e. a. Beringer, *Review of Particle Physics*, Phys.Rev. D **86** (2012) 010001.

[122] L. C. Andrews, *Special functions of mathematics for engineers*, SPIE Optical Engineering Press - Oxford University Press (1998), (pag. 110).

[123] LHCb Collaboration, R. Aaij et al., *Measurement of the fragmentation fraction ratio f_s/f_d and its dependence on B meson kinematics*, JHEP 1304 (2013) 001, arXiv:1301.5286 [hep-ex].

[124] R. Brun, P. Canal, and F. Rademakers, *Design, Development and Evolution of the ROOT System*, ACAT2010, PoS **002** (2010).

[125] T. Junk, *Confidence Level Computation for Combining Searches with Small Statistics*, Nucl. Instrum. Meth. A **434** (1999), 435-443, arXiv:hep-ex/9902006.

[126] A. L. Read, *Presentation of search results: the CL_s technique*, J. Phys. G: Nucl. Part. Phys **28** (2002), 2693-2704.

[127] K. De Bruyn, R. Fleischer, R. Knegjens, P. Koppenburg, M. Merk, et al., *Probing New Physics via the $B_s^0 \to \mu^+\mu^-$ Effective Lifetime* Phys.Rev.Lett. 109 (2012) 041801, arXiv:1204.1737 [hep-ph].

[128] E. Ritsch, *Fast Calorimeter Punch-Through Simulation for the ATLAS Experiment*, CERN-THESIS-2011-112.

[129] ATLAS Collaboration, *Luminosity Determination in pp Collisions at $\sqrt{s} = 7$ TeV Using the ATLAS Detector at the LHC*, Eur. Phys. J. C **71** (2011) 1630, arXiv:1101.2185 [hep-ex].

[130] A. W. F. Edwards, *Likelihood*, (1992), ISBN-13: 978-0801844430.

[131] R. Barlow, *Extended maximum likelihood*, Nuclear Instruments and Methods in Physics Research A **297** (1990) 496.

[132] G. Cowan, *Statistical Data Analysis*, Clarendon Press, Oxford (1998), ISBN: 0-19-850155-8.

[133] K. De Bruyn et al., *Probing New Physics via the $B_s^0 \to \mu^+\mu^-$ Effective Lifetime*, Phys. Rev. Lett. **109** (2012) 041801, arXiv:1204.1737 [hep-ph].

[134] W. Verkerke, D. Kirkby, *The RooFit toolkit for data modeling*, arXiv:physics/0306116 [physics.data-an].

[135] Particle Data Group Collaboration, *2013 Review of Particle Physics*, `http://pdg.lbl.gov/2013/reviews/rpp2013-rev-probability.pdf`.

[136] R. A. Fisher, *The Use of Multiple Measures in Taxonomic Problems*, Annals of Eugenics **7** (1936) 179-188.

[137] K.A. Olive et al. (Particle Data Group), Chin. Phys. C **38** 090001 (2014).

[138] G. C. Fox and S. Wolfram., *Observables for the Analysis of Event Shapes in e^+e^- Annihilation and Other Processes*, Phys. Rev. Lett. **41** (1978), 1851.

[139] T. Sjöstrand et al, *High-Energy-Physics Event Generation with PYTHIA 6.1*, Computer Phys. Commun. **135** (2001) 238, arXiv:hep-ph/0010017.

[140] D. J. Lange, *The EvtGen particle decay simulation package*, Nucl. Instr. Meth. A **462** (2001) 152.

[141] M. Dobbs, J. Hansen, *The HepMC C++ Monte Carlo event record for High Energy Physics*, Comput. Phys. Commun. **134** (2001) 41.

[142] The ATLAS Collaboration (ATLAS Computing Group), *ATLAS Computing Technical Design Report*, ATLAS-TDR-017, CERN-LHCC-2005-022 (2005).

[143] G. Barrand et al., *Gaudi - A Software Architecture and Framework for building HEP Data Processing Applications*, International Conference on Computing in High Energy Physics (CHEP) (2000).

[144] M. Smizanska, *PythiaB: an interface to Pythia 6 dedicated to simulation of beauty events*, ATL-COM-PHYS-2003-038, 2003.

[145] N. L. Johnson, *Systems of Frequency Curves Generated by Methods of Translation*, Biometrika, **36** (December 1949), 149-76.

[146] C. Jones and A. Pewsey, *Sinh-arcsinh distributions: a broad family giving rise to powerful tests of normality and symmetry*, Tech. Rep. 08/06 (2008), The Open University.

[147] K.A. Olive et al. (Particle Data Group), Chin. Phys. C38, 090001 (2014).